

10/520060

0715 Rec'd PCT/PTO 30 DEC 2004

Express Mail Label No.

Dated: _____

Docket No.: 04270/0202279-US0
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of:
Christopher B. Phelps et al.

Application No.: Not Yet Assigned

Confirmation No.: N/A

Filed: Concurrently Herewith

Art Unit: N/A

For: **LIGAND BINDING DOMAINS OF NUCLEAR
HORMONE RECEPTORS**

Examiner: Not Yet Assigned

AFFIRMATION OF CLAIM FOR PRIORITY

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

Applicant hereby claims priority under 35 U.S.C. 119 based on the following prior foreign application filed in the following foreign country on the date indicated:

Country	Application No.	Date
United Kingdom	0215295.7	July 2, 2002

In support of this claim, attached is Form PCT/IB/304 evidencing receipt of the priority document on August 8, 2003 during prosecution of International Application No. PCT/GB2003/002844.

Dated: December 30, 2004

Respectfully submitted,

By  *Form 1820
(53,970)*
Paul F. Fehlner, Ph.D.

Registration No.: 35,135
DARBY & DARBY P.C.
P.O. Box 5257
New York, New York 10150-5257
(212) 527-7700
(212) 753-6237 (Fax)
Attorneys/Agents For Applicant

{W:\04270\0202279us0\00331010.DOC [REDACTED]}

BEST AVAILABLE COPY



INVESTOR IN PEOPLE

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

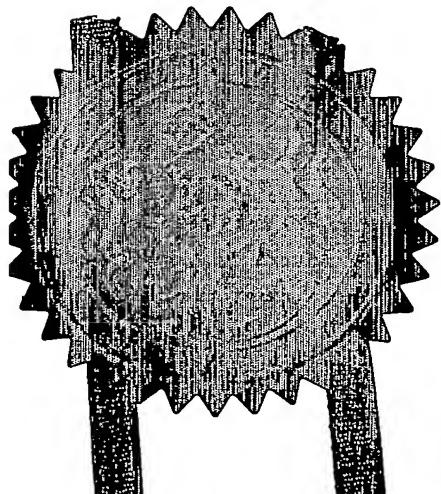
REC'D 08 AUG 2003
WIPO PCT

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

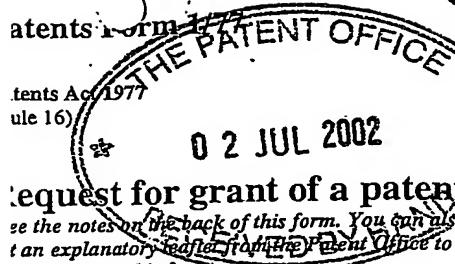
In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.



Signed

Dated 22 July 2003



The
Patent
Office

02 JUL 02 E 030403 A 0300137
PO 17700 0-01-02 15295.7

The Patent Office

Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference

P031003GB

2. Patent application number
(The Patent Office will fill in this part)

0215295.7

3. Full name, address and postcode of the or of
each applicant (*underline all surnames*)

INPHARMATICA LIMITED
60 Charlotte Street.
London
W1T 2NU

7853286062

Patents ADP number (*if you know it*)

If the applicant is a corporate body, give the
country/state of its incorporation

United Kingdom

4. Title of the invention

PROTEINS

5. Name of your agent (*if you have one*)

Carpmaels & Ransford

"Address for service" in the United Kingdom
to which all correspondence should be sent
(including the postcode)

43 Bloomsbury Square
London
WC1A 2RA

Patents ADP number (*if you know it*)

83001

6. If you are declaring priority from one or more
earlier patent applications, give the country
and the date of filing of the or of each of these
earlier applications and (*if you know it*) the or
each application number

Country

Priority application number
(if you know it)

Date of filing
(day / month / year)

7. If this application is divided or otherwise
derived from an earlier UK application,
give the number and the filing date of
the earlier application

Number of earlier application

Date of filing
(day / month / year)

8. Is a statement of inventorship and of right
to grant of a patent required in support of
this request? (Answer 'Yes' if:

- a) any applicant named in part 3 is not an inventor, or
- b) there is an inventor who is not named as an
applicant, or
- c) any named applicant is a corporate body

Yes

See note (d))

9. Enter the number of sheets for any of the following items you are filing with this form.
Do not count copies of the same document

Continuation sheets of this form

Description	88
Claim(s)	10
Abstract	1
Drawing(s)	43 + 43

[Handwritten mark]

10. If you are also filing any of the following, state how many against each item.

Priority documents

Translations of priority documents

Statement of inventorship and right to grant of a patent (*Patents Form 7/77*)

Request for preliminary examination and search (*Patents Form 9/77*)

Request for substantive examination (*Patents Form 10/77*)

Any other documents
(please specify)

11. I/We request the grant of a patent on the basis of this application.

Signature

Date

Carpmaels & Ransford
Carpmaels & Ransford 2nd July 2002

12. Name and daytime telephone number of person to contact in the United Kingdom

Hugh Goodfellow

020-7242 8692

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

- If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.
- Write your answers in capital letters using black ink or you may type them.
- If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.
- If you have answered 'Yes' Patents Form 7/77 will need to be filed.
- Once you have filled in the form you must remember to sign and date it.
- For details of the fee and ways to pay please contact the Patent Office.

PROTEINS

This invention relates to a protein, termed CAA05410.2 and to functional equivalents of this protein, herein identified as containing novel Nuclear Hormone Receptor Ligand Binding Domains and to the use of these proteins and nucleic acid sequence from the 5 encoding genes in the diagnosis, prevention and treatment of disease.

All publications, patents and patent applications cited herein are incorporated in full by reference.

BACKGROUND

The process of drug discovery is presently undergoing a fundamental revolution as the 10 era of functional genomics comes of age. The term "functional genomics" applies to an approach utilising bioinformatics tools to ascribe function to protein sequences of interest. Such tools are becoming increasingly necessary as the speed of generation of sequence data is rapidly outpacing the ability of research laboratories to assign functions to these protein sequences.

15 As bioinformatics tools increase in potency and in accuracy, these tools are rapidly replacing the conventional techniques of biochemical characterisation. Indeed, the advanced bioinformatics tools used in identifying the present invention are now capable of outputting results in which a high degree of confidence can be placed.

Various institutions and commercial organisations are examining sequence data as they 20 become available and significant discoveries are being made on an on-going basis. However, there remains a continuing need to identify and characterise further genes and the polypeptides that they encode, as targets for research and for drug discovery.

Recently, a remarkable tool for the evaluation of sequences of unknown function has 25 been developed by the Applicant for the present invention. This tool is a database system, termed the Biopendium search database, that is the subject of co-pending International Patent Application No. PCT/GB01/01105. This database system consists of an integrated data resource created using proprietary technology and containing information generated from an all-by-all comparison of all available protein or nucleic acid sequences.

The aim behind the integration of these sequence data from separate data resources is to

combine as much data as possible, relating both to the sequences themselves and to information relevant to each sequence, into one integrated resource. All the available data relating to each sequence, including data on the three-dimensional structure of the encoded protein, if this is available, are integrated together to make best use of the 5 information that is known about each sequence and thus to allow the most educated predictions to be made from comparisons of these sequences. The annotation that is generated in the database and which accompanies each sequence entry imparts a biologically relevant context to the sequence information.

This data resource has made possible the accurate prediction of protein function from 10 sequence alone. Using conventional technology, this is only possible for proteins that exhibit a high degree of sequence identity (above about 20%-30% identity) to other proteins in the same functional family. Accurate predictions are not possible for proteins that exhibit a very low degree of sequence homology to other related proteins of known function.

15 In the present case, a protein whose sequence is recorded in a publicly available database as CAA05410.2 (NCBI Genebank nucleotide accession number AJ002425.2 and a Genebank protein accession number CAA05410.2), is implicated as containing a novel Nuclear Hormone Receptor Ligand Binding Domain.

Introduction to Nuclear Hormone Receptor Ligand Binding Domains

20 The Nuclear Hormone Receptor superfamily (see Table 1) encodes structurally related proteins that regulate the transcription of target genes. These proteins include receptors for steroid and thyroid hormones, vitamins, and other proteins for which no ligands have been found. To be classified as a "Nuclear Hormone Receptor" a protein must possess at least one of two key domains; a C4-type zinc finger DNA-Binding Domain (DBD) or a 25 Ligand Binding Domain (LBD). The DBD is required for binding DNA in the vicinity of target genes, and the LBD is required for steroid-like ligand responsiveness. It is the Ligand Binding Domain of Nuclear Hormone Receptors which is the binding site for pharmacological agents such as Tamoxifen.

Many Nuclear Hormone receptors possess both a DBD and an LBD, and a well-known 30 example of this is Estrogen receptor alpha, which possesses both a DBD and an LBD.

There are also members of the Nuclear Hormone Receptor family which posses a DBD but lack an LBD; for example the proteins Knirps (SWISS-PROT code P10734) and ODR7 (SWISS-PROT code P41933) possess DBDs, but both lack LBDs. Implicit in the existence of proteins such as Knirps and ODR7 is the fact that possession of a DBD does 5 not mean that a LBD will be concomitantly present.

There are also members of the Nuclear Hormone Receptor family that possess an LBD but lack a DBD; for example the protein "Short Heterodimer Partner", SHP (SWISS-PROT code Q15466) possesses an LBD but lacks a DBD.

A further refinement in the classification of Nuclear Hormone Receptors is to classify on 10 the basis of possession of an LBD. Nuclear Hormone receptors which possess an LBD can be sub-classified as "Nuclear Hormone Receptor Ligand Binding Domain" family members. Thus Estrogen receptor alpha and SHP are "Nuclear Hormone Receptor Ligand Binding Domain" family members whereas Knirps and ODR7 are excluded.

The DBD directs the protein to bind specific DNA sequences in the vicinity of target 15 genes.

The Ligand Binding Domain (LBD) binds and responds to the cognate hormone. Ligand binding to the LBD triggers a conformational change which expels a bound "Nuclear Receptor Co-Repressor". The site previously occupied by the Co-Repressor is then free to recruit a "Nuclear Receptor Co-Activator". This Ligand-triggered swap of a Co-Repressor for a Co-Activator is the mechanism by which Ligand binding leads to the 20 transcriptional activation of target genes. The LBD is the binding site for all Nuclear Hormone Receptor targeted drugs to date and it is thus desirable to identify novel Ligand Binding Domains since these will be attractive drug targets. The LBD also directs dimerisation with other LBDs. For example the Estrogen receptor alpha ligand binding 25 domain can homodimerise with itself, or heterodimerise with the Estrogen receptor beta ligand binding domain. Ligand Binding Domains share low sequence identity (~15%) but have very similar structures and so present ideal targets for a structure-based relationship tool such as Genome Threader.

Table 1: Nuclear Hormone Receptor Superfamily

Class ID	Nuclear Hormone Receptor Name	Species	Accession
NR1 group			
NR1A1	Thyroid Hormone Receptor alpha	HUMAN	M24748
NR1A2	Thyroid Hormone Receptor beta	HUMAN	X04707
NR1A3	Thyroid Hormone Receptor Ciona	CIONA	AF077403
NR1B1	Retinoic Acid Receptor alpha	HUMAN	X06538
NR1B2	Retinoic Acid Receptor beta	HUMAN	Y00291
NR1B3	Retinoic Acid Receptor gamma	HUMAN	M57707
NR1B4	Retinoic Acid Receptor Polyandrocarpa	POLYANDROCARPA	D86615
NR1C1	Peroxisome Proliferator Activated Receptor alpha	HUMAN	L02932
NR1C2	Peroxisome Proliferator Activated Receptor beta	HUMAN	L07592
NR1C3	Peroxisome Proliferator Activated Receptor gamma	HUMAN	L40904
NR1D1	Rev-erbA	HUMAN	M24898
NR1D2	Rev-erbB	HUMAN	L31785
NR1D3	E75	FLY	X51548
NR1E1	E78	FLY	U01087
NR1F1	RAR-related Orphan Receptor alpha	HUMAN	U04897
NR1F2	RAR-related Orphan Receptor beta	HUMAN	Y08639
NR1F3	RAR-related Orphan Receptor gamma	HUMAN	U16997
NR1F4	DHR3	FLY	M90806
NR1G1	CNR14	WORM	U13074
NR1H1	Ecdysone Receptor	FLY	M74078
NR1H2	Liver X Receptor beta	HUMAN	U07132
NR1H3	Liver X Receptor alpha	HUMAN	U22662
NR1H4	Farnesoid X Receptor	HUMAN	U68233
NR1I1	Vitamin D Receptor	HUMAN	J03258
NR1I2	Pregnane X Receptor	HUMAN	AF061056
NR1I3	Constitutive Androstane Receptor alpha	HUMAN	Z30425
NR1I4	Constitutive Androstane Receptor beta	MOUSE	AF009327
NR1J1	DHR96	FLY	U36792
NR1K1	NHR1	WORM	U19360
NR2 group			
NR2A1	Hepatocyte Nuclear Factor 4 alpha	HUMAN	X76930
NR2A2	Hepatocyte Nuclear Factor 4 beta	XENOPUS	Z49827
NR2A3	Hepatocyte Nuclear Factor 4 gamma	HUMAN	Z49826
NR2A4	Drosophila Hepatocyte Nuclear Factor 4	FLY	U70874
NR2B1	Retinoid X Receptor alpha	HUMAN	X52773
NR2B2	Retinoid X Receptor beta	HUMAN	M84820
NR2B3	Retinoid X Receptor gamma	HUMAN	U38480
NR2B4	Ultraspiracle	FLY	X53417
NR2C1	TR2	HUMAN	M29960
NR2C2	TR4	HUMAN	L27586
NR2D1	SpSHR2	SEARUCHIN	U38281
NR2E1	TLX	HUMAN	Y13276
NR2E2	Tailless	FLY	AF019362
NR2E3	Photoreceptor-specific Nuclear Receptor	HUMAN	AF121129

NR2E4	Dissatisfaction	FLY	AF106677
NR2E5	FAX-1	WORM	AF176087
NR2F1	COUP-TFI	HUMAN	X12795
NR2F2	COUP-TFII	HUMAN	M64497
NR2F3	Seven-up	FLY	M28863
NR2F4	Xenopus COUP-TFIII	XENOPUS	X63092
NR2F5	Zebrafish COUP-TFIII	ZEBRAFISH	X70300
NR2F6	EAR2	HUMAN	X12794
NR3 group			
NR3A1	Estrogen Receptor alpha	HUMAN	P03372
NR3A2	Estrogen Receptor beta	HUMAN	AB006590
NR3B1	Estrogen Receptor Related alpha	HUMAN	X51416
NR3B2	Estrogen Receptor Related beta	HUMAN	AF094517
NR3B3	Estrogen Receptor Related gamma	HUMAN	AF058291
NR3C1	Glucocorticoid Receptor	HUMAN	X03225
NR3C2	Mineralocorticoid Receptor	HUMAN	M16801
NR3C3	Progesterone Receptor	HUMAN	M15716
NR3C4	Androgen Receptor	HUMAN	M20132
NR4 group			
NR4A1	NGFI-Balpha	HUMAN	L13740
NR4A2	NGFI-Bbeta	HUMAN	X75918
NR4A3	NGFI-Bgamma	HUMAN	D78579
NR4A4	DHR38	FLY	X89246
NR5 group			
NR5A1	FTZ-F1	HUMAN	U76388
NR5A2	FTF	HUMAN	U93553
NR5A3	Drosophila FTZ-F1	FLY	M98397
NR5A4	Zebrafish FTZ-F1	ZEBRAFISH	AF198086
NR5B1	FTZ-F1B	FLY	L06423
NR6 group			
NR6A1	Germ Cell Nuclear Factor	HUMAN	U64876
NR6A2	GCNF Related Factor	TENEBRI	AF124981
NR0 group (have only one characteristic domain)			
NR0A1 sub-group (have only DBD no LBD)			
NR0A1	Knirps	FLY	X13331
NR0A2	Knirps-related	FLY	X14153
NR0A3	Embryonic gonad	FLY	X16631
NR0A4	ODR7	WORM	U16708
NR0B sub-group (have only LBD no DBD)			
NR0B1	DAX1	HUMAN	S74720
NR0B2	Short Heterodimer Partner SHP	HUMAN	L76571

Table data taken from Laudet and Gronemeyer "The Nuclear Receptor Facts Book" Academic Press. ClassID refers to a classification code for each member, and accession refers to NCBI GenBank nucleotide accession code.

II. Nuclear Hormone Receptor Ligand Binding Domain Family and Disease

Nuclear Hormone Receptor Ligand Binding Domain family members have been shown to play a role in diverse physiological functions, many of which can play a role in disease processes (see Table 2).

5

Table 2. Nuclear Hormone Receptors and disease.

Nuclear Hormone Receptor	Disease
Androgen Receptor	Androgen Insensitivity Syndrome (Lubahn <i>et al.</i> 1989 Proc. Natl. Acad. Sci. USA 86, 9534-9538).
	Reifenstein syndrome (Wooster <i>et al.</i> 1992 Nat. Genet. 2, 132-134).
	X-linked recessive spinal and bulbar muscular atrophy (MacLean <i>et al.</i> 1995 Mol. Cell. Endocrinol. 112, 133-141).
	Male breast cancer ((Wooster <i>et al.</i> 1992 Nat. Genet. 2, 132-134).
Glucocorticoid Receptor	Nelson's syndrome (Karl <i>et al.</i> 1996 J. Clin. Endocrinol. Metab. 81, 124-129).
	Glucocorticoid resistant acute T-cell leukemia (Hala <i>et al.</i> 1996 Int. J. Cancer 68, 663-668)..
Mineralocorticoid Receptor	Pseudohypoaldosteronism (Chung <i>et al.</i> 1995 J. Clin. Endocrinol. Metab. 80, 3341-3345).
Estrogen Receptor alpha and beta	ER alpha expression is elevated in a subset of human breast cancers. The application of Tamoxifen is the major therapy to prevent breast tumour progression. (Petrangeli <i>et al.</i> 1994 J. Steroid Biochem. Mol. Biol. 49, 327-331). Estrogen Receptors are involved in cancer, particular cancers originating from estrogen-responsive tissues, including breast (as mentioned above), uterus and prostate, myeloproliferative disorders, such as leukemia, hypertension, hypotension, fertility enhancement, contraception, pregnancy termination, progesterone antagonism, wound healing, scarring, obesity, dermatological disorders including cellulite, estrogen-mediated hair characteristics, central nervous system disorders, Alzheimer's disease, cognition enhancement, learning and memory enhancement, immunomodulation, and osteoporosis (taken from Lauden and Gronemeyer, "The Nuclear Receptor Facts Book" Academic Press).

Vitamin D3 Receptor	Mutations in the Vitamin D3 receptor produce a hereditary disorder similar in phenotype to Vitamin D3 deficiency (Rickets) (Hughes <i>et al.</i> 1988 <i>Science</i> 242, 1702-1725).
Retinoic Acid Receptor alpha	Acute Myeloid Leukemia (Lavau and Dejean 1994 <i>Leukemia</i> 8, 9-15).
Thyroid Hormone Receptor beta	“Generalised Resistance to Thyroid Hormones” (GRTH) (Refetoff 1994 <i>Thyroid</i> 4, 345-349).
DAX1	X-linked Adrenal Hypoplasia Congenita (AHC) and Hypogonadism (Ito <i>et al.</i> 1997 <i>Mol. Cell. Biol.</i> 17, 1476-1483).

Alteration of Nuclear Hormone Receptor Ligand Binding Domain family members by ligands which bind to their LBD thus provides a means to alter the disease phenotype. There is thus a great need for the identification of novel Nuclear Hormone Receptor Ligand Binding Domains, as these proteins may play a role in the diseases identified above, as well as in other disease states. The identification of novel Nuclear Hormone Receptor Ligand Binding Domains is thus highly relevant for the treatment and diagnosis of disease, particularly those identified in Table 2.

THE INVENTION

The invention is based on the discovery that a region of the CAA05410.2 protein functions as a Nuclear Hormone Receptor Ligand Binding Domain.

For the CAA05410.2 protein, it has been found that a region including residues 394-604 of this protein sequence adopts an equivalent fold to residues 15 to 247 (equivalent to residues 320 to 552 in the full-length numbering scheme) of the Human Estrogen Receptor alpha Ligand Binding Domain (PDB code 3ERT:A). Human Estrogen Receptor alpha Ligand Binding Domain is known to function as a Nuclear Hormone Receptor Ligand Binding Domain. This relationship is not just to the Human Estrogen Receptor alpha Ligand Binding Domain, but rather to the Nuclear Hormone Receptor Ligand Binding Domain family as a whole.

The discovery of sharing an equivalent fold allows the functional annotation of this region of CAA05410.2, and therefore proteins that include this region, as possessing Nuclear Hormone Receptor Ligand Binding Domain activity.

In a first aspect, the invention provides a polypeptide, which polypeptide:

- (i) comprises the amino acid sequence as recited in SEQ ID NO:2;
- (ii) is a fragment thereof having Nuclear Hormone Receptor Ligand Binding Domain activity or having an antigenic determinant in common with the polypeptides of 5 (i); or
- (iii) is a functional equivalent of (i) or (ii).

Preferably, a polypeptide according to this aspect of the invention consists of the amino acid sequence as recited in SEQ ID NO:2, or is a fragment or functional equivalent thereof.

10 The polypeptide having the sequence recited in SEQ ID NO:2 is referred to hereafter as "the LBDG5 polypeptide".

According to this aspect of the invention, a preferred polypeptide fragment according to part ii) above includes the region of the LBDG5 polypeptide that is predicted as that responsible for Nuclear Hormone Receptor Ligand Binding Domain activity (hereafter, the "LBDG5

15 Nuclear Hormone Receptor Ligand Binding Domain region"), or is a variant thereof. As defined herein, the LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region is considered to extend between residue 394 and residue 604 of the LBDG5 polypeptide sequence.

According to a further aspect of this invention, a preferred functional equivalent according

20 to part iii) above is a polypeptide comprising the amino acid sequence recited in SEQ ID No:4. This polypeptide, herein termed LBDG10, is a *Rattus norvegicus* orthologue of the LBDG5 polypeptide. On the basis of the high homology to CAA05410.2 (LBDG5), we annotate CAA05409.2 (LBDG10) as also containing a Nuclear Hormone Receptor Ligand Binding Domain. In addition to this homology, the LBDG10 polypeptide has Genome

25 ThreaderTM and Inpharmatica PSI-BLAST relationships which consolidate the annotation of the polypeptide as containing a Nuclear Hormone Receptor Ligand Binding Domain. The region of the LBDG10 polypeptide that has been identified as containing this domain (hereafter, the "LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region") is a region from residue 413 to residue 603 of the polypeptide. This domain adopts an equivalent

fold to the Human Estrogen Receptor alpha Ligand Binding Domain.

The LBDG10 polypeptide sequence is recorded in publicly available databases as CAA05409.2 (NCBI Genbank nucleotide accession number AJ002424.2 and Genbank protein accession number CAA05409.2), but has so far remained unannotated functionally
5 as containing a Nuclear Hormone Receptor Ligand Binding Domain.

According to a further aspect of this invention, a preferred functional equivalent according to part iii) above is a polypeptide comprising the amino acid sequence recited in SEQ ID No:6. This polypeptide, herein termed LBDG11, is a *Homo sapiens* parologue of the LBDG5 polypeptide. On the basis of the high homology to CAA05410.2 (LBDG5), we
10 annotate BAB62888.1 (LBDG11) as also containing a Nuclear Hormone Receptor Ligand Binding Domain. The region of the LBDG11 polypeptide that has been identified as containing this domain (hereafter, the "LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region") is a region from residue 118 to residue 319 of the polypeptide.

The LBDG11 polypeptide sequence is recorded in publicly available databases as
15 BAB62888.1 (NCBI Genbank nucleotide accession number AB035676.1 and a Genbank protein accession number BAB62888.1), but has so far remained unannotated functionally.

According to a further aspect of this invention, a preferred functional equivalent according to part iii) above is a polypeptide comprising the amino acid sequence recited in SEQ ID No:8. This polypeptide, herein termed LBDG12, is a *Mus musculus* orthologue of the LBDG11 polypeptide. On the basis of the high homology to CAA05410.2 (LBDG5), we
20 annotate AAH03486.1 (LBDG12) as also containing a Nuclear Hormone Receptor Ligand Binding Domain. The region of the LBDG12 polypeptide that has been identified as containing this domain (hereafter, the "LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region") is a region from residue 118 to residue 318 of the polypeptide.

25 The LBDG12 polypeptide sequence is recorded in publicly available databases as AAH03486.1 (NCBI Genbank nucleotide accession number BC003486.1 and a Genbank protein accession number AAH03486.1), but has so far remained unannotated functionally.

According to a further aspect of this invention, a preferred functional equivalent according to part iii) above is a polypeptide comprising the amino acid sequence recited in SEQ ID

No:10. This polypeptide, herein termed LBDG13, is a *Rattus norvegicus* orthologue of the LBDG11 polypeptide. On the basis of the high homology to CAA05410.2 (LBDG5), we annotate AAK49953.1 (LBDG13) as also containing a Nuclear Hormone Receptor Ligand Binding Domain. The region of the LBDG13 polypeptide that has been identified as 5 containing this domain (hereafter, the "LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region") is a region from residue 117 to residue 317 of the polypeptide.

The LBDG13 polypeptide sequence is recorded in publicly available databases as AAK49953.1 (NCBI Genbank nucleotide accession number AF348701.1 and a Genbank protein accession number AAK49953.1), but has so far remained unannotated functionally.

10 This aspect of the invention also includes fusion proteins that incorporate polypeptide fragments and variants of these polypeptide fragments as defined above, provided that said fusion proteins possess activity as a Nuclear Hormone Receptor Ligand Binding Domain.

It is considered highly likely that the polypeptides and polypeptide fragments as defined above will in action form homodimers and also heterodimers with other proteins containing

15 Nuclear Hormone Receptor Ligand Binding Domains. Accordingly, a second embodiment of the first aspect of the invention provides a LBDG5, LBDG10, LBDG11, LBDG12 or LBDG13 polypeptide or polypeptide fragment as these terms are defined in i), ii) or iii) above in the form of a dimer complex. For the LBDG5 and LBDG10 polypeptides, examples of Nuclear Hormone Receptor Ligand Binding Domain partners include the 20 Estrogen receptor α (ER α) and Estrogen receptor β (ER β). Accordingly, a further embodiment of the first aspect of the invention provides a LBDG5 and LBDG10 polypeptide or polypeptide fragment as these terms are defined in i), ii) or iii) above in the form of a dimer complex with the Estrogen receptor α (ER α) and Estrogen receptor β (ER β).

25 Such a dimer complex may be in the form of a heterodimer complex, said complex comprising a LBDG5, LBDG10, LBDG11, LBDG12 or LBDG13 polypeptide or polypeptide fragment as these terms are defined in i) or ii) above, wherein the polypeptide partner in the heterodimer complex is selected from the broad group of Nuclear Hormone Receptor Ligand Binding Domains. In the case of the LBDG5 and LBDG10 polypeptides, examples of Nuclear Hormone Receptor Ligand Binding Domain 30 binding partners include the human estrogen receptor α ligand binding domain and the

estrogen receptor β ligand binding domain. In the case of the LBDG10 polypeptide, preferred polypeptide partners include Rat estrogen receptor α ligand binding domain and the Rat estrogen receptor β ligand binding domain.

In a second aspect, the invention provides a purified nucleic acid molecule that encodes a polypeptide or a dimer complex of the first aspect of the invention. Preferably, the purified nucleic acid molecule has the nucleic acid sequence as recited in SEQ ID NO:1 (encoding the LBDG5 polypeptide), the nucleic acid sequence as recited in SEQ ID NO:3 (encoding the LBDG10 polypeptide), the nucleic acid sequence as recited in SEQ ID NO:5 (encoding the LBDG11 polypeptide), the nucleic acid sequence as recited in SEQ ID NO:7 (encoding the LBDG12 polypeptide), the nucleic acid sequence as recited in SEQ ID NO:9 (encoding the LBDG13 polypeptide), or is a redundant equivalent or fragment of either of these sequences. A preferred nucleic acid fragment is one that encodes a polypeptide fragment according to part ii) above, preferably a polypeptide fragment that includes the LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region, a polypeptide fragment that includes the LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region, a polypeptide fragment that includes the LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region, a polypeptide fragment that includes the LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region, a polypeptide fragment that includes the LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region or that encodes a variant of these fragments as this term is defined above.

To encode a homo- or heterodimer complex comprising the polypeptide of SEQ ID NO:2, or the polypeptide of SEQ ID NO:2, a nucleic acid molecule should incorporate the coding sequence for both this polypeptide and for its binding partner, optionally connected by a linker sequence. However, the components of the dimer need not be encoded by the same nucleic acid molecule, but may simply be co-expressed together under conditions allowing the two components to dimerise through their mutual affinity for one another.

In a third aspect, the invention provides a purified nucleic acid molecule which hybridizes under high stringency conditions with a nucleic acid molecule of the second

aspect of the invention.

In a fourth aspect, the invention provides a vector, such as an expression vector, that contains a nucleic acid molecule of the second or third aspect of the invention.

In a fifth aspect, the invention provides a host cell transformed with a vector of the fourth
5 aspect of the invention.

In a sixth aspect, the invention provides a ligand which binds specifically to, and which preferably inhibits the Nuclear Hormone Receptor Ligand Binding Domain activity of, a polypeptide of the first aspect of the invention.

In a seventh aspect, the invention provides a compound that is effective to alter the
10 expression of a natural gene which encodes a polypeptide of the first aspect of the invention or to regulate the activity of a polypeptide of the first aspect of the invention.

A compound of the seventh aspect of the invention may either increase (agonise) or decrease (antagonise) the level of expression of the gene or the activity of the polypeptide. Importantly, the identification of the function of the region defined herein as
15 the LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG5 polypeptide, the LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG10 polypeptide, the LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG11 polypeptide, the LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG12 polypeptide, and the LBDG13 Nuclear
20 Hormone Receptor Ligand Binding Domain region of the LBDG13 polypeptide respectively, allows for the design of screening methods capable of identifying compounds that are effective in the treatment and/or diagnosis of diseases in which Nuclear Hormone Receptor Ligand Binding Domains are implicated. Ligands and compounds according to the sixth and seventh aspects of the invention may be identified
25 using such methods. These methods are included as aspects of the present invention.

In an eighth aspect, the invention provides a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a ligand of the fifth aspect of the invention, or a compound of the sixth aspect of the invention, for use in therapy or diagnosis of a

disease in which Nuclear Hormone Receptor Ligand Binding Domains are implicated. These molecules may also be used in the manufacture of a medicament for the treatment of such diseases. Examples of diseases in which Nuclear Hormone Receptor Ligand Binding Domains are implicated include cell proliferative disorders, including neoplasm, 5 melanoma, lung, colorectal, breast, uterus, prostate, pancreas, head and neck and other solid tumours, myeloproliferative disorders, such as leukemia, non-Hodgkin lymphoma, leukopenia, thrombocytopenia, angiogenesis disorder, Kaposi's sarcoma, autoimmune/inflammatory disorders, including allergy, inflammatory bowel disease, arthritis, psoriasis and respiratory tract inflammation, asthma, and organ transplant 10 rejection, cardiovascular disorders, including hypertension, hypotension, oedema, angina, atherosclerosis, thrombosis, sepsis, shock, reperfusion injury, heart arrhythmia, and ischemia, neurological disorders including, central nervous system disease, Alzheimer's disease, Parkinson's disease, brain injury, stroke, amyotrophic lateral sclerosis, anxiety, depression, and pain, cognition enhancement, learning and memory enhancement, 15 developmental disorders, metabolic disorders including diabetes mellitus, osteoporosis, lipid metabolism disorder, hyperthyroidism, hyperparathyroidism, thyroid hormone resistance syndrome, hypercalcemia, hypocalcaemia, hypercholesterolemia, hyperlipidemia, and obesity, renal disorders, including glomerulonephritis, renovascular hypertension, blood disorders including hemophilia, dermatological disorders, including, 20 cellulite, acne, eczema, psoriasis and wound healing, scarring, negative effects of aging, fertility enhancement, contraception, pregnancy termination, progesterone antagonism, hormone replacement therapies, steroid hormone-like mediated hair characteristics, immunomodulation, AIDS, vision disorders, glucocorticoid resistance, mineralocorticoid resistance, androgen resistance, pseudohypoaldosteronism, spinal/bulbar muscular 25 atrophy, extraskeletal myxoid chondrosarcomas, adrenal insufficiency, sexual reversal, infections including viral infection, bacterial infection, fungal infection and parasitic infection and other pathological conditions, particularly those in which nuclear hormone receptors are implicated.

The prediction that the LBDG5 polypeptide and its functional equivalent, the LBDG10 30 polypeptide form heterodimers with polypeptides that contain Nuclear Hormone Receptor Ligand Binding Domains, particularly ER α and ER β , also makes this

polypeptide, the LBDG10 polypeptide and compounds of the nature described above that bind it of potential utility in the diagnosis and treatment of diseases in which ER α and ER β are implicated. Such diseases include cancer, particular cancers originating from estrogen-responsive tissues, including breast, uterus and prostate, myeloproliferative

5 disorders, such as leukemia, hypertension, hypotension, fertility enhancement, contraception, pregnancy termination, progesterone antagonism, wound healing, scarring, obesity, dermatological disorders including cellulite, estrogen-mediated hair characteristics, central nervous system disorders, Alzheimer's disease, cognition enhancement, learning and memory enhancement, immunomodulation, osteoporosis,

10 among other pathological conditions.

In a ninth aspect, the invention provides a method of diagnosing a disease in a patient, comprising assessing the level of expression of a natural gene encoding a polypeptide of the first aspect of the invention or the activity of a polypeptide of the first aspect of the invention in tissue from said patient and comparing said level of expression or activity to a control level, wherein a level that is different to said control level is indicative of disease. Such a method will preferably be carried out *in vitro*. Similar methods may be used for monitoring the therapeutic treatment of disease in a patient, wherein altering the level of expression or activity of a polypeptide or nucleic acid molecule over the period of time towards a control level is indicative of regression of disease.

20 A preferred method for detecting polypeptides of the first aspect of the invention comprises the steps of: (a) contacting a ligand, such as an antibody, of the sixth aspect of the invention with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and (b) detecting said complex.

A number of different such methods according to the ninth aspect of the invention exist, 25 as the skilled reader will be aware, such as methods of nucleic acid hybridization with short probes, point mutation analysis, polymerase chain reaction (PCR) amplification and methods using antibodies to detect aberrant protein levels. Similar methods may be used on a short or long term basis to allow therapeutic treatment of a disease to be monitored in a patient. The invention also provides kits that are useful in these methods for 30 diagnosing disease.

In a tenth aspect, the invention provides for the use of a polypeptide of the first aspect of the invention as a Nuclear Hormone Receptor Ligand Binding Domain. The invention also provides for the use of a nucleic acid molecule according to the second or third aspects of the invention to express a protein that possesses Nuclear Hormone Receptor 5 Ligand Binding Domain activity. The invention also provides a method for effecting Nuclear Hormone Receptor Ligand Binding Domain activity, said method utilising a polypeptide of the first aspect of the invention.

In an eleventh aspect, the invention provides a pharmaceutical composition comprising a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second 10 or third aspect of the invention, or a vector of the fourth aspect of the invention, or a host cell of the fifth aspect of the invention, or a ligand of the sixth aspect of the invention, or a compound of the seventh aspect of the invention, in conjunction with a pharmaceutically-acceptable carrier.

In a twelfth aspect, the present invention provides a polypeptide of the first aspect of the 15 invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a host cell of the fifth aspect of the invention, or a ligand of the sixth aspect of the invention, or a compound of the seventh aspect of the invention, for use in the manufacture of a medicament for the diagnosis or treatment of a disease in which Nuclear Hormone Receptor Ligand Binding Domains are 20 implicated, examples of which include cell proliferative disorders, including neoplasm, melanoma, lung, colorectal, breast, uterus, prostate, pancreas, head and neck and other solid tumours, myeloproliferative disorders, such as leukemia, non-Hodgkin lymphoma, leukopenia, thrombocytopenia, angiogenesis disorder, Kaposi's sarcoma, autoimmune/inflammatory disorders, including allergy, inflammatory bowel disease, 25 arthritis, psoriasis and respiratory tract inflammation, asthma, and organ transplant rejection, cardiovascular disorders, including hypertension, hypotension, oedema, angina, atherosclerosis, thrombosis, sepsis, shock, reperfusion injury, heart arrhythmia, and ischemia, neurological disorders including, central nervous system disease, Alzheimer's disease, Parkinson's disease, brain injury, stroke, amyotrophic lateral sclerosis, anxiety, 30 depression, and pain, cognition enhancement, learning and memory enhancement, developmental disorders, metabolic disorders including diabetes mellitus, osteoporosis,

lipid metabolism disorder, hyperthyroidism, hyperparathyroidism, thyroid hormone resistance syndrome, hypercalcemia, hypocalcaemia, hypercholesterolemia, hyperlipidemia, and obesity, renal disorders, including glomerulonephritis, renovascular hypertension, blood disorders including hemophilia, dermatological disorders, including, 5 cellulite, acne, eczema, psoriasis and wound healing, scarring, negative effects of aging, fertility enhancement, contraception, pregnancy termination, progesterone antagonism, hormone replacement therapies, steroid hormone-like mediated hair characteristics, immunomodulation, AIDS, vision disorders, glucocorticoid resistance, mineralocorticoid resistance, androgen resistance, pseudohypoaldosteronism, spinal/bulbar muscular 10 atrophy, extraskeletal myxoid chondrosarcomas, adrenal insufficiency, sexual reversal, infections including viral infection, bacterial infection, fungal infection and parasitic infection and other pathological conditions, particularly those in which nuclear hormone receptors are implicated.

In a thirteenth aspect, the invention provides a method of treating a disease in a patient 15 comprising administering to the patient a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a ligand of the sixth aspect of the invention, or a compound of the seventh aspect of the invention.

For diseases in which the expression of a natural gene encoding a polypeptide of the first 20 aspect of the invention, or in which the activity of a polypeptide of the first aspect of the invention, is lower in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, ligand or compound administered to the patient should be an agonist. Conversely, for diseases in which the expression of the natural gene or activity of the polypeptide is higher in a diseased patient 25 when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, ligand or compound administered to the patient should be an antagonist. Examples of such antagonists include antisense nucleic acid molecules, ribozymes and ligands, such as antibodies.

In a fourteenth aspect, the invention provides transgenic or knockout non-human animals 30 that have been transformed to express higher, lower or absent levels of a polypeptide of

the first aspect of the invention. Such transgenic animals are very useful models for the study of disease and may also be used in screening regimes for the identification of compounds that are effective in the treatment or diagnosis of such a disease.

5 A summary of standard techniques and procedures which may be employed in order to utilise the invention is given below. It will be understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors and reagents described. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only and it is not intended that this terminology should limit the scope of the present invention. The extent of the invention is limited only by the terms of
10 the appended claims.

Standard abbreviations for nucleotides and amino acids are used in this specification.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA technology and immunology, which are within the skill of those working in the art.

15 Such techniques are explained fully in the literature. Examples of particularly suitable texts for consultation include the following: Sambrook Molecular Cloning; A Laboratory Manual, Second Edition (1989); DNA Cloning, Volumes I and II (D.N. Glover ed. 1985); Oligonucleotide Synthesis (M.J. Gait ed. 1984); Nucleic Acid Hybridization (B.D. Hames & S.J. Higgins eds. 1984); Transcription and Translation (B.D. Hames & S.J.
20 Higgins eds. 1984); Animal Cell Culture (R.I. Freshney ed. 1986); Immobilized Cells and Enzymes (IRL Press, 1986); B. Perbal, A Practical Guide to Molecular Cloning (1984); the Methods in Enzymology series (Academic Press, Inc.), especially volumes 154 & 155; Gene Transfer Vectors for Mammalian Cells (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Immunochemical Methods in Cell and Molecular
25 Biology (Mayer and Walker, eds. 1987, Academic Press, London); Scopes, (1987) Protein Purification: Principles and Practice, Second Edition (Springer Verlag, N.Y.); and Handbook of Experimental Immunology, Volumes I-IV (D.M. Weir and C. C. Blackwell eds. 1986).

As used herein, the term "polypeptide" includes any peptide or protein comprising two or
30 more amino acids joined to each other by peptide bonds or modified peptide bonds, i.e.

peptide isosteres. This term refers both to short chains (peptides and oligopeptides) and to longer chains (proteins).

The polypeptide of the present invention may be in the form of a mature protein or may be a pre-, pro- or prepro- protein that can be activated by cleavage of the pre-, pro- or 5 prepro- portion to produce an active mature polypeptide. In such polypeptides, the pre-, pro- or prepro- sequence may be a leader or secretory sequence or may be a sequence that is employed for purification of the mature polypeptide sequence.

The polypeptide of the first aspect of the invention may form part of a fusion protein. For example, it is often advantageous to include one or more additional amino acid sequences 10 which may contain secretory or leader sequences, pro-sequences, sequences which aid in purification, or sequences that confer higher protein stability, for example during recombinant production. Alternatively or additionally, the mature polypeptide may be fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol).

15 Polypeptides may contain amino acids other than the 20 gene-encoded amino acids, modified either by natural processes, such as by post-translational processing or by chemical modification techniques which are well known in the art. Among the known modifications which may commonly be present in polypeptides of the present invention are glycosylation, lipid attachment, sulphation, gamma-carboxylation, for instance of glutamic acid residues, hydroxylation and ADP-ribosylation. Other potential 20 modifications include acetylation, acylation, amidation, covalent attachment of flavin, covalent attachment of a haeme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulphide bond formation, 25 demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, GPI anchor formation, iodination, methylation, myristylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

30 Modifications can occur anywhere in a polypeptide, including the peptide backbone, the

amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl terminus in a polypeptide, or both, by a covalent modification is common in naturally-occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention.

5 The modifications that occur in a polypeptide often will be a function of how the polypeptide is made. For polypeptides that are made recombinantly, the nature and extent of the modifications in large part will be determined by the post-translational modification capacity of the particular host cell and the modification signals that are present in the amino acid sequence of the polypeptide in question. For instance,
10 glycosylation patterns vary between different types of host cell.

The polypeptides of the present invention can be prepared in any suitable manner. Such polypeptides include isolated naturally-occurring polypeptides (for example purified from cell culture), recombinantly-produced polypeptides (including fusion proteins), synthetically-produced polypeptides or polypeptides that are produced by a combination
15 of these methods.

The functionally-equivalent polypeptides of the first aspect of the invention may be polypeptides that are homologous to the LBDG5 polypeptide. Two polypeptides are said to be "homologous", as the term is used herein, if the sequence of one of the polypeptides has a high enough degree of identity or similarity to the sequence of the other
20 polypeptide. "Identity" indicates that at any particular position in the aligned sequences, the amino acid residue is identical between the sequences. "Similarity" indicates that, at any particular position in the aligned sequences, the amino acid residue is of a similar type between the sequences. Degrees of identity and similarity can be readily calculated (Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press, New
25 York, 1988; Biocomputing. Informatics and Genome Projects, Smith, D.W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part 1, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991).
30 Homologous polypeptides therefore include natural biological variants (for example,

allelic variants or geographical variations within the species from which the polypeptides are derived) and mutants (such as mutants containing amino acid substitutions, insertions or deletions) of the LBDG5 polypeptide. Such mutants may include polypeptides in which one or more of the amino acid residues are substituted with a conserved or non-
5 conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code. Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; among the basic residues Lys and Arg; or among the aromatic residues Phe and Tyr. Particularly preferred are variants in
10 which several, i.e. between 5 and 10, 1 and 5, 1 and 3, 1 and 2 or just 1 amino acids are substituted, deleted or added in any combination. Especially preferred are silent substitutions, additions and deletions, which do not alter the properties and activities of the protein. Also especially preferred in this regard are conservative substitutions. Such mutants also include polypeptides in which one or more of the amino acid residues
15 includes a substituent group.

Typically, greater than 30% identity between two polypeptides (preferably, over a specified region) is considered to be an indication of functional equivalence. Preferably, functionally equivalent polypeptides of the first aspect of the invention have a degree of sequence identity with the LBDG5 polypeptide, or with active fragments thereof, of
20 greater than 80%. More preferred polypeptides have degrees of identity of greater than 85%, 90%, 95%, 98% or 99%, respectively with the LBDG5 polypeptide, or with active fragments thereof.

Percentage identity, as referred to herein, is as determined using BLAST version 2.1.3 using the default parameters specified by the NCBI (the National Center for
25 Biotechnology Information; <http://www.ncbi.nlm.nih.gov/>) [Blosum 62 matrix; gap open penalty=11 and gap extension penalty=1].

Preferred functionally-equivalent homologues of the LBDG5 polypeptide are the LBDG10 polypeptide, the LBDG11 polypeptide, the LBDG12 polypeptide and the LBDG13 polypeptide referred to above. The LBDG10 polypeptide has 83% sequence
30 identity with the LBDG5 polypeptide.

In the present case, preferred active fragments of the LBDG5 polypeptide are those that include the LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region. As discussed above, the LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region is considered to extend between residue 394 and residue 604 of the LBDG5 polypeptide 5 sequence.

Preferred active fragments of the LBDG10 polypeptide are those that include the LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region. As discussed above, the LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region is considered to extend between residue 413 and residue 603 of the LBDG10 polypeptide 10 sequence.

Preferred active fragments of the LBDG11 polypeptide are those that include the LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region. As discussed above, the LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region is considered to extend between residue 118 and residue 319 of the LBDG11 polypeptide 15 sequence.

Preferred active fragments of the LBDG12 polypeptide are those that include the LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region. As discussed above, the LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region is considered to extend between residue 118 and residue 318 of the LBDG12 polypeptide 20 sequence.

Preferred active fragments of the LBDG13 polypeptide are those that include the LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region. As discussed above, the LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region is considered to extend between residue 117 and residue 317 of the LBDG13 polypeptide 25 sequence.

The functionally-equivalent polypeptides of the first aspect of the invention may also be polypeptides which have been identified using one or more techniques of structural alignment. For example, the Inpharmatica Genome ThreaderTM technology that forms one aspect of the search tools used to generate the Biopendium search database may be used 30 (see co-pending International patent application PCT/GB01/01105) to identify

polypeptides of presently-unknown function which, while having low sequence identity as compared to the LBDG5 polypeptide, are predicted to have Nuclear Hormone Receptor Ligand Binding Domain activity, by virtue of sharing significant structural homology with the LBDG5 polypeptide sequence.

5 By "significant structural homology" is meant that the Inpharmatica Genome ThreaderTM predicts two proteins, or protein regions, to share structural homology with a certainty of at least 10% more preferably, at least 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and above. The certainty value of the Inpharmatica Genome ThreaderTM is calculated as follows. A set of comparisons was initially performed using the Inpharmatica Genome

10 ThreaderTM exclusively using sequences of known structure. Some of the comparisons were between proteins that were known to be related (on the basis of structure). A neural network was then trained on the basis that it needed to best distinguish between the known relationships and known not-relationships taken from the CATH structure classification (www.biochem.ucl.ac.uk/bsm/cath). This resulted in a neural network score

15 between 0 and 1. However, again as the number of proteins that are related and the number that are unrelated were known, it was possible to partition the neural network results into packets and calculate empirically the percentage of the results that were correct. In this manner, any genuine prediction in the Biopendium search database has an attached neural network score and the percentage confidence is a reflection of how

20 successful the Inpharmatica Genome ThreaderTM was in the training/testing set.

Structural homologues of LBDG5 should share structural homology with the LBDG5 Nuclear Hormone Receptor Ligand Binding Domain. Such structural homologues are predicted to have Nuclear Hormone Receptor Ligand Binding Domain activity by virtue of sharing significant structural homology with this polypeptide sequence.

25 The polypeptides of the first aspect of the invention also include fragments of the LBDG5 polypeptide, functional equivalents of the fragments of the LBDG5 polypeptide, and fragments of the functional equivalents of the LBDG5 polypeptides (such as fragments of the LBDG10 polypeptide, the LBDG11 polypeptide, the LBDG12 polypeptide, and the LBDG13 polypeptide), provided that those functional equivalents and fragments retain

30 Nuclear Hormone Receptor Ligand Binding Domain activity or have an antigenic

determinant in common with the LBDG5 polypeptide.

As used herein, the term "fragment" refers to a polypeptide having an amino acid sequence that is the same as part, but not all, of the amino acid sequence of the LBDG5 polypeptides or one of its functional equivalents. The fragments should comprise at least 5 n consecutive amino acids from the sequence and, depending on the particular sequence, n preferably is 7 or more (for example, 8, 10, 12, 14, 16, 18, 20 or more). Small fragments may form an antigenic determinant.

Preferred polypeptide fragments according to this aspect of the invention are fragments that include a region defined herein as the LBDG5 Nuclear Hormone Receptor Ligand 10 Binding Domain region of the LBDG5 polypeptide, the LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG10 polypeptide, the LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG11 polypeptide, the LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG12 polypeptide, and the LBDG13 Nuclear Hormone Receptor Ligand Binding Domain 15 region of the LBDG13 polypeptide respectively. These regions are the regions that have been annotated as Nuclear Hormone Receptor Ligand Binding Domains.

For the LBDG5 polypeptide, this region is considered to extend between residue 394 and residue 604. For the LBDG10 polypeptide, this region is considered to extend between residue 413 and residue 603. For the LBDG11 polypeptide, this region is considered to 20 extend between residue 118 and residue 319. For the LBDG12 polypeptide, this region is considered to extend between residue 118 and residue 318. For the LBDG13 polypeptide, this region is considered to extend between residue 117 and residue 317.

Variants of these fragments are included as embodiments of this aspect of the invention, provided that these variants possess activity as a Nuclear Hormone Receptor 25 Ligand Binding Domain.

In one respect, the term "variant" is meant to include extended or truncated versions of this polypeptide fragment.

For extended variants, it is considered highly likely that the Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG5 polypeptide, the LBDG10 polypeptide, the

LBDG11 polypeptide, the LBDG12 polypeptide, or the LBDG13 polypeptide will fold correctly and show Nuclear Hormone Receptor Ligand Binding Domain activity if additional residues C terminal and/or N terminal of these boundaries in the LBDG5 polypeptide sequence, the LBDG10 polypeptide sequence, the LBDG11 polypeptide sequence, the LBDG12 polypeptide sequence, or the LBDG13 polypeptide sequence respectively are included in the polypeptide fragment. For example, an additional 5, 10, 20, 30, 40 or even 50 or more amino acid residues may be included at either or both the C terminal and/or N terminal of the boundaries of the Nuclear Hormone Receptor Ligand Binding Domain regions of the polypeptide, without prejudicing the ability of the polypeptide fragment to fold correctly and exhibit Nuclear Hormone Receptor Ligand Binding Domain activity.

For truncated variants of the LBDG5 polypeptide, the LBDG10 polypeptide, the LBDG11 polypeptide, the LBDG12 polypeptide, or the LBDG13 polypeptide, one or more amino acid residues may be deleted at either or both the C terminus or the N terminus of the Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG5 polypeptide, the LBDG10 polypeptide, the LBDG11 polypeptide, the LBDG12 polypeptide, or the LBDG13 polypeptide, respectively.

In a second respect, the term "variant" includes homologues of the polypeptide fragments described above, that possess significant sequence homology with the Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG5 polypeptide or one of its functional equivalents, provided that said variants retain activity as an Nuclear Hormone Receptor Ligand Binding Domain.

Homologues include those polypeptide molecules that possess greater than 30% identity with the LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region, the LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region, the LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region, the LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region, or the LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region. Percentage identity is as determined using BLAST version 2.1.3 using the default parameters specified by the NCBI (the National Center for Biotechnology Information; <http://www.ncbi.nlm.nih.gov/>) [Blosum 62

matrix; gap open penalty=11 and gap extension penalty=1]. Preferably, variant homologues of polypeptide fragments of this aspect of the invention have a degree of sequence identity with the LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region, the LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region, the 5 LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region, the LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region, or the LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region, of greater than 40%. More preferred variant polypeptides have degrees of identity of greater than 50%, 60%, 70%, 80%, 90%, 95%, 98% or 99%, respectively with these polypeptide regions, provided that said 10 variants retain activity as a Nuclear Hormone Receptor Ligand Binding Domain. Variant polypeptides also include homologues of the truncated forms of the polypeptide fragments discussed above, provided that said variants retain activity as a Nuclear Hormone Receptor Ligand Binding Domain.

The polypeptide fragments of the first aspect of the invention may be polypeptide fragments that exhibit significant structural homology with the structure of the 15 polypeptide fragment defined by the LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region, the LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region, the LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region, the LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region, or the LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region, for example, as identified by 20 the Inpharmatica Genome Threader™. Accordingly, polypeptide fragments that are structural homologues of the polypeptide fragments defined by LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region, the LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region, the LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region, the LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region, or the LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region should adopt the same fold as that adopted by this polypeptide fragment, 25 as this fold is defined above.

Such fragments may be "free-standing", i.e. not part of or fused to other amino acids or 30 polypeptides, or they may be comprised within a larger polypeptide of which they form a part or region. When comprised within a larger polypeptide, the fragment of the invention

most preferably forms a single continuous region. For instance, certain preferred embodiments relate to a fragment having a pre- and/or pro- polypeptide region fused to the amino terminus of the fragment and/or an additional region fused to the carboxyl terminus of the fragment. However, several fragments may be comprised within a single 5 larger polypeptide.

Homo and heterodimers of the polypeptide fragments and variants as these are described above are also included in this aspect of the invention. Preferred polypeptide partners in heterodimers of the LBDG5 and LBDG10 polypeptides include polypeptides that contain Nuclear Hormone Receptor Ligand Binding Domains, such as ER α and ER β .

10 The polypeptides of the present invention or their immunogenic fragments (comprising at least one antigenic determinant) can be used to generate ligands, such as polyclonal or monoclonal antibodies, that are immunospecific for the polypeptides. Such antibodies may be employed to isolate or to identify clones expressing the polypeptides of the invention or to purify the polypeptides by affinity chromatography. The antibodies may 15 also be employed as diagnostic or therapeutic aids, amongst other applications, as will be apparent to the skilled reader.

The term "immunospecific" means that the antibodies have substantially greater affinity for the polypeptides of the invention than their affinity for other related polypeptides in the prior art. As used herein, the term "antibody" refers to intact molecules as well as to 20 fragments thereof, such as Fab, F(ab') 2 and Fv, which are capable of binding to the antigenic determinant in question. Such antibodies thus bind to the polypeptides of the first aspect of the invention.

If polyclonal antibodies are desired, a selected mammal, such as a mouse, rabbit, goat or horse, may be immunised with a polypeptide of the first aspect of the invention. The 25 polypeptide used to immunise the animal can be derived by recombinant DNA technology or can be synthesized chemically. If desired, the polypeptide can be conjugated to a carrier protein. Commonly used carriers to which the polypeptides may be chemically coupled include bovine serum albumin, thyroglobulin and keyhole limpet haemocyanin. The coupled polypeptide is then used to immunise the animal. Serum from

the immunised animal is collected and treated according to known procedures, for example by immunoaffinity chromatography.

Monoclonal antibodies to the polypeptides of the first aspect of the invention can also be readily produced by one skilled in the art. The general methodology for making 5 monoclonal antibodies using hybridoma technology is well known (see, for example, Kohler, G. and Milstein, C., *Nature* 256: 495-497 (1975); Kozbor et al., *Immunology Today* 4: 72 (1983); Cole et al., 77-96 in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc. (1985)).

Panels of monoclonal antibodies produced against the polypeptides of the first aspect of 10 the invention can be screened for various properties, i.e., for isotype, epitope, affinity, etc. Monoclonal antibodies are particularly useful in purification of the individual polypeptides against which they are directed. Alternatively, genes encoding the monoclonal antibodies of interest may be isolated from hybridomas, for instance by PCR techniques known in the art, and cloned and expressed in appropriate vectors.

15 Chimeric antibodies, in which non-human variable regions are joined or fused to human constant regions (see, for example, Liu et al., *Proc. Natl. Acad. Sci. USA*, 84, 3439 (1987)), may also be of use.

The antibody may be modified to make it less immunogenic in an individual, for example 20 by humanisation (see Jones et al., *Nature*, 321, 522 (1986); Verhoeyen et al., *Science*, 239: 1534 (1988); Kabat et al., *J. Immunol.*, 147: 1709 (1991); Queen et al., *Proc. Natl Acad. Sci. USA*, 86, 10029 (1989); Gorman et al., *Proc. Natl Acad. Sci. USA*, 88: 34181 (1991); and Hodgson et al., *Bio/Technology* 9: 421 (1991)). The term "humanised 25 antibody", as used herein, refers to antibody molecules in which the CDR amino acids and selected other amino acids in the variable domains of the heavy and/or light chains of a non-human donor antibody have been substituted in place of the equivalent amino acids in a human antibody. The humanised antibody thus closely resembles a human antibody but has the binding ability of the donor antibody.

In a further alternative, the antibody may be a "bispecific" antibody, that is an antibody 30 having two different antigen binding domains, each domain being directed against a different epitope.

Phage display technology may be utilised to select genes which encode antibodies with binding activities towards the polypeptides of the invention either from repertoires of PCR amplified V-genes of lymphocytes from humans screened for possessing the relevant antibodies, or from naive libraries (McCafferty, J. *et al.*, (1990), *Nature* 348, 5 552-554; Marks, J. *et al.*, (1992) *Biotechnology* 10, 779-783). The affinity of these antibodies can also be improved by chain shuffling (Clackson, T. *et al.*, (1991) *Nature* 352, 624-628).

Antibodies generated by the above techniques, whether polyclonal or monoclonal, have additional utility in that they may be employed as reagents in immunoassays, 10 radioimmunoassays (RIA) or enzyme-linked immunosorbent assays (ELISA). In these applications, the antibodies can be labelled with an analytically-detectable reagent such as a radioisotope, a fluorescent molecule or an enzyme.

Preferred nucleic acid molecules of the second and third aspects of the invention are those which encode the polypeptide sequences recited in SEQ ID NO:2 (particularly the 15 coding region, nucleotides 61-1905), SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, or SEQ ID NO:10 and functionally equivalent polypeptides, including active fragments of the LBDG5 polypeptide, the LBDG10 polypeptide, the LBDG11 polypeptide, the LBDG12 polypeptide, and the LBDG13 polypeptide such as a fragment including the LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG5 20 polypeptide sequence, a fragment including the LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG10 polypeptide sequence, a fragment including the LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG11 polypeptide sequence, a fragment including the LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG12 polypeptide sequence, a 25 fragment including the LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG13 polypeptide sequence or a homologue thereof and homo- and heterodimers of the LBDG5 polypeptide and the LBDG10 polypeptide and/or its active fragments.

Nucleic acid molecules encompassing these stretches of sequence form a preferred 30 embodiment of this aspect of the invention.

These nucleic acid molecules may be used in the methods and applications described herein. The nucleic acid molecules of the invention preferably comprise at least n consecutive nucleotides from the sequences disclosed herein where, depending on the particular sequence, n is 10 or more (for example, 12, 14, 15, 18, 20, 25, 30, 35, 40 or 5 more).

The nucleic acid molecules of the invention also include sequences that are complementary to nucleic acid molecules described above (for example, for antisense or probing purposes).

Nucleic acid molecules of the present invention may be in the form of RNA, such as 10 mRNA, or in the form of DNA, including, for instance cDNA, synthetic DNA or genomic DNA. Such nucleic acid molecules may be obtained by cloning, by chemical synthetic techniques or by a combination thereof. The nucleic acid molecules can be prepared, for example, by chemical synthesis using techniques such as solid phase 15 phosphoramidite chemical synthesis, from genomic or cDNA libraries or by separation from an organism. RNA molecules may generally be generated by the *in vitro* or *in vivo* transcription of DNA sequences.

The nucleic acid molecules may be double-stranded or single-stranded. Single-stranded DNA may be the coding strand, also known as the sense strand, or it may be the non-coding strand, also referred to as the anti-sense strand.

20 The term "nucleic acid molecule" also includes analogues of DNA and RNA, such as those containing modified backbones, and peptide nucleic acids (PNA). The term "PNA", as used herein, refers to an antisense molecule or an anti-gene agent which comprises an oligonucleotide of at least five nucleotides in length linked to a peptide backbone of amino acid residues, which preferably ends in lysine. The terminal lysine confers 25 solubility to the composition. PNAs may be pegylated to extend their lifespan in a cell, where they preferentially bind complementary single stranded DNA and RNA and stop transcript elongation (Nielsen, P.E. *et al.* (1993) *Anticancer Drug Des.* 8:53-63).

A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:2, or an active fragment thereof, may be identical to the coding sequence of the nucleic acid molecule 30 shown in SEQ ID NO:1. A nucleic acid molecule which encodes the polypeptide of SEQ

ID NO:4, or an active fragment thereof, may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:3. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:6, or an active fragment thereof, may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:5. A nucleic acid
5 molecule which encodes the polypeptide of SEQ ID NO:8, or an active fragment thereof, may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:7. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:10, or an active fragment thereof, may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:9. These molecules also may have a different sequence
10 which, as a result of the degeneracy of the genetic code, encodes the polypeptide SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, or an active fragment of the LBDG5 polypeptide, the LBDG10 polypeptide, the LBDG11 polypeptide, the LBDG12 polypeptide, or the LBDG13 polypeptide, such as a fragment including the LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region, the
15 LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region, the LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region, the LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region, the LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region, or a homologue thereof. The LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region is considered to extend between
20 residue 394 and residue 604 of the LBDG5 polypeptide sequence. In SEQ ID NO:1 the LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region is thus encoded by a nucleic acid molecule including nucleotide 1240 to nucleotide 1872. The LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region is considered to extend between residue 413 and residue 603 of the LBDG10 polypeptide sequence. In SEQ ID
25 NO:3 the LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region is thus encoded by a nucleic acid molecule including nucleotide 1237 to nucleotide 1809. The LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region is considered to extend between residue 118 and residue 319 of the LBDG11 polypeptide sequence. In SEQ ID NO:5 the LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region
30 is thus encoded by a nucleic acid molecule including nucleotide 352 to nucleotide 957. The LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region is considered

to extend between residue 118 and residue 318 of the LBDG12 polypeptide sequence. In SEQ ID NO:7 the LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region is thus encoded by a nucleic acid molecule including nucleotide 352 to nucleotide 954. The LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region is considered 5 to extend between residue 117 and residue 317 of the LBDG13 polypeptide sequence. In SEQ ID NO:9 the LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region is thus encoded by a nucleic acid molecule including nucleotide 349 to nucleotide 951. Nucleic acid molecules encompassing these stretches of sequence, and homologues of these sequences, form a preferred embodiment of this aspect of the invention.

10 Such nucleic acid molecules that encode the polypeptide of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8 or SEQ ID NO:10 may include, but are not limited to, the coding sequence for the mature polypeptide by itself; the coding sequence for the mature polypeptide and additional coding sequences, such as those encoding a leader or secretory sequence, such as a pro-, pre- or prepro- polypeptide sequence; the coding 15 sequence of the mature polypeptide, with or without the aforementioned additional coding sequences, together with further additional, non-coding sequences, including non-coding 5' and 3' sequences, such as the transcribed, non-translated sequences that play a role in transcription (including termination signals), ribosome binding and mRNA stability. The nucleic acid molecules may also include additional sequences which encode 20 additional amino acids, such as those which provide additional functionalities.

The nucleic acid molecules of the second and third aspects of the invention may also encode the fragments or the functional equivalents of the polypeptides and fragments of the first aspect of the invention.

As discussed above, a preferred fragment of the LBDG5 polypeptide is a fragment 25 including the LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region, or a homologue thereof. The Nuclear Hormone Receptor Ligand Binding Domain region is encoded by a nucleic acid molecule including nucleotide 1240 to nucleotide 1872 of SEQ ID NO:1. A preferred fragment of the LBDG10 polypeptide is a fragment including the LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region, or a homologue 30 thereof. The LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region is

encoded by a nucleic acid molecule including nucleotide 1237 to nucleotide 1809 of SEQ ID NO:3. A preferred fragment of the LBDG11 polypeptide is a fragment including the LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region, or a homologue thereof. The LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region is 5 encoded by a nucleic acid molecule including nucleotide 352 to nucleotide 957 of SEQ ID NO:5. A preferred fragment of the LBDG12 polypeptide is a fragment including the LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region, or a homologue thereof. The LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region is encoded by a nucleic acid molecule including nucleotide 352 to nucleotide 954 of SEQ 10 ID NO:7. A preferred fragment of the LBDG13 polypeptide is a fragment including the LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region, or a homologue thereof. The LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region is encoded by a nucleic acid molecule including nucleotide 349 to nucleotide 951 of SEQ ID NO:9.

15 Functionally equivalent nucleic acid molecules according to the invention may be naturally-occurring variants such as a naturally-occurring allelic variant, or the molecules may be a variant that is not known to occur naturally. Such non-naturally occurring variants of the nucleic acid molecule may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells or organisms.

20 Among variants in this regard are variants that differ from the aforementioned nucleic acid molecules by nucleotide substitutions, deletions or insertions. The substitutions, deletions or insertions may involve one or more nucleotides. The variants may be altered in coding or non-coding regions or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or insertions.

25 The nucleic acid molecules of the invention can also be engineered, using methods generally known in the art, for a variety of reasons, including modifying the cloning, processing, and/or expression of the gene product (the polypeptide). DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides are included as techniques which may be used to engineer the nucleotide 30 sequences. Site-directed mutagenesis may be used to insert new restriction sites, alter

glycosylation patterns, change codon preference, produce splice variants, introduce mutations and so forth.

Nucleic acid molecules which encode a polypeptide of the first aspect of the invention may be ligated to a heterologous sequence so that the combined nucleic acid molecule 5 encodes a fusion protein. Such combined nucleic acid molecules are included within the second or third aspects of the invention. For example, to screen peptide libraries for inhibitors of the activity of the polypeptide, it may be useful to express, using such a combined nucleic acid molecule, a fusion protein that can be recognised by a commercially-available antibody. A fusion protein may also be engineered to contain a 10 cleavage site located between the sequence of the polypeptide of the invention and the sequence of a heterologous protein so that the polypeptide may be cleaved and purified away from the heterologous protein.

The nucleic acid molecules of the invention also include antisense molecules that are partially complementary to nucleic acid molecules encoding polypeptides of the present 15 invention and that therefore hybridize to the encoding nucleic acid molecules (hybridization). Such antisense molecules, such as oligonucleotides, can be designed to recognise, specifically bind to and prevent transcription of a target nucleic acid encoding a polypeptide of the invention, as will be known by those of ordinary skill in the art (see, for example, Cohen, J.S., Trends in Pharm. Sci., 10, 435 (1989), Okano, J. Neurochem. 20 56, 560 (1991); O'Connor, J. Neurochem 56, 560 (1991); Lee *et al.*, Nucleic Acids Res 6, 3073 (1979); Cooney *et al.*, Science 241, 456 (1988); Dervan *et al.*, Science 251, 1360 (1991).

The term "hybridization" as used here refers to the association of two nucleic acid 25 molecules with one another by hydrogen bonding. Typically, one molecule will be fixed to a solid support and the other will be free in solution. Then, the two molecules may be placed in contact with one another under conditions that favour hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment 30 of the liquid phase molecule to the solid support (Denhardt's reagent or BLOTO); the concentration of the molecules; use of compounds to increase the rate of association of

molecules (dextran sulphate or polyethylene glycol); and the stringency of the washing conditions following hybridization (see Sambrook *et al.* [*supra*]).

The inhibition of hybridization of a completely complementary molecule to a target molecule may be examined using a hybridization assay, as known in the art (see, for example, Sambrook *et al* [*supra*]). A substantially homologous molecule will then compete for and inhibit the binding of a completely homologous molecule to the target molecule under various conditions of stringency, as taught in Wahl, G.M. and S.L. Berger (1987; *Methods Enzymol.* 152:399-407) and Kimmel, A.R. (1987; *Methods Enzymol.* 152:507-511).

"Stringency" refers to conditions in a hybridization reaction that favour the association of very similar molecules over association of molecules that differ. High stringency hybridisation conditions are defined as overnight incubation at 42°C in a solution comprising 50% formamide, 5XSSC (150mM NaCl, 15mM trisodium citrate), 50mM sodium phosphate (pH7.6), 5x Denhardts solution, 10% dextran sulphate, and 20 microgram/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1X SSC at approximately 65°C. Low stringency conditions involve the hybridisation reaction being carried out at 35°C (see Sambrook *et al.* [*supra*]). Preferably, the conditions used for hybridization are those of high stringency.

Preferred embodiments of this aspect of the invention are nucleic acid molecules that are at least 70% identical over their entire length to a nucleic acid molecule encoding the LBDG5 polypeptide (SEQ ID NO:2), the LBDG10 polypeptide (SEQ ID NO:4), the LBDG11 polypeptide (SEQ ID NO:6), the LBDG12 polypeptide (SEQ ID NO:8), the LBDG13 polypeptide (SEQ ID NO:10), and nucleic acid molecules that are substantially complementary to such nucleic acid molecules. A preferred active fragment is a fragment that includes an LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG5 polypeptide sequence. A further preferred active fragment is a fragment that includes an LBDG10 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG10 polypeptide sequence. A further preferred active fragment is a fragment that includes an LBDG11 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG11 polypeptide sequence. A further preferred active fragment is a fragment that

includes an LBDG12 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG12 polypeptide sequence. A further preferred active fragment is a fragment that includes an LBDG13 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG13 polypeptide sequence. Accordingly, preferred nucleic acid molecules include 5 those that are at least 70% identical over their entire length to a nucleic acid molecule encoding the Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG5 polypeptide sequence, the LBDG10 polypeptide sequence, the LBDG11 polypeptide sequence, the LBDG12 polypeptide sequence, and the LBDG13 polypeptide sequence.

Percentage identity, as referred to herein, is as determined using BLAST version 2.1.3 10 using the default parameters specified by the NCBI (the National Center for Biotechnology Information; <http://www.ncbi.nlm.nih.gov/>).

Preferably, a nucleic acid molecule according to this aspect of the invention comprises a region that is at least 80% identical over its entire length to the nucleic acid molecule having the sequence given in SEQ ID NO:1, to a region including nucleotides 1240 to 15 1872 of SEQ ID NO:1, the nucleic acid molecule having the sequence given in SEQ ID NO:3, to a region including nucleotides 1237 to 1809 of SEQ ID NO:3, the nucleic acid molecule having the sequence given in SEQ ID NO:5, to a region including nucleotides 352 to 957 of SEQ ID NO:5, the nucleic acid molecule having the sequence given in SEQ ID NO:7, to a region including nucleotides 352 to 954 of SEQ ID NO:7, the nucleic acid 20 molecule having the sequence given in SEQ ID NO:9, to a region including nucleotides 349-951 of SEQ ID NO:9, or a nucleic acid molecule that is complementary to any one of these regions of nucleic acid. In this regard, nucleic acid molecules at least 90%, preferably at least 95%, more preferably at least 98% or 99% identical over their entire length to the same are particularly preferred. Preferred embodiments in this respect are 25 nucleic acid molecules that encode polypeptides which retain substantially the same biological function or activity as the LBDG5 polypeptide, the LBDG10 polypeptide, the LBDG11 polypeptide, the LBDG12 polypeptide, or the LBDG13 polypeptide.

The invention also provides a process for detecting a nucleic acid molecule of the invention, comprising the steps of: (a) contacting a nucleic probe according to the 30 invention with a biological sample under hybridizing conditions to form duplexes; and

(b) detecting any such duplexes that are formed.

As discussed additionally below in connection with assays that may be utilised according to the invention, a nucleic acid molecule as described above may be used as a hybridization probe for RNA, cDNA or genomic DNA, in order to isolate full-length cDNAs and genomic clones encoding the LBDG5 polypeptide or its functional equivalents and to isolate cDNA and genomic clones of homologous or orthologous genes that have a high sequence similarity to the gene encoding this polypeptide.

In this regard, the following techniques, among others known in the art, may be utilised and are discussed below for purposes of illustration. Methods for DNA sequencing and analysis are well known and are generally available in the art and may, indeed, be used to practice many of the embodiments of the invention discussed herein. Such methods may employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase (US Biochemical Corp, Cleveland, OH), Taq polymerase (Perkin Elmer), thermostable T7 polymerase (Amersham, Chicago, IL), or combinations of polymerases and proof-reading exonucleases such as those found in the ELONGASE Amplification System marketed by Gibco/BRL (Gaithersburg, MD). Preferably, the sequencing process may be automated using machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno, NV), the Peltier Thermal Cycler (PTC200; MJ Research, Watertown, MA) and the ABI Catalyst and 373 and 377 DNA Sequencers (Perkin Elmer).

One method for isolating a nucleic acid molecule encoding a polypeptide with an equivalent function to that of the LBDG5 polypeptide, particularly with an equivalent function to the LBDG5 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG5 polypeptide, is to probe a genomic or cDNA library with a natural or artificially-designed probe using standard procedures that are recognised in the art (see, for example, "Current Protocols in Molecular Biology", Ausubel *et al.* (eds). Greene Publishing Association and John Wiley Interscience, New York, 1989,1992). Probes comprising at least 15, preferably at least 30, and more preferably at least 50, contiguous bases that correspond to, or are complementary to, nucleic acid sequences from the appropriate encoding gene are particularly useful probes. In the case of SEQ ID NO:1, particularly suitable probes may be selected from nucleotides 1240 to 1872. In the case of SEQ ID

NO:3, particularly suitable probes may be selected from nucleotides 1237 to 1809. In the case of SEQ ID NO:5, particularly suitable probes may be selected from nucleotides 352 to 957. In the case of SEQ ID NO:7, particularly suitable probes may be selected from nucleotides 352 to 954. In the case of SEQ ID NO:9, particularly suitable probes may be selected from nucleotides 349 to 951.

Such probes may be labelled with an analytically-detectable reagent to facilitate their identification. Useful reagents include, but are not limited to, radioisotopes, fluorescent dyes and enzymes that are capable of catalysing the formation of a detectable product. Using these probes, the ordinarily skilled artisan will be capable of isolating complementary copies of genomic DNA, cDNA or RNA polynucleotides encoding proteins of interest from human, mammalian or other animal sources and screening such sources for related sequences, for example, for additional members of the family, type and/or subtype.

In many cases, isolated cDNA sequences will be incomplete, in that the region encoding the polypeptide will be cut short, normally at the 5' end. Several methods are available to obtain full length cDNAs, or to extend short cDNAs. Such sequences may be extended utilising a partial nucleotide sequence and employing various methods known in the art to detect upstream sequences such as promoters and regulatory elements. For example, one method which may be employed is based on the method of Rapid Amplification of cDNA Ends (RACE; see, for example, Frohman et al., Proc. Natl. Acad. Sci. USA (1988) 85: 8998-9002). Recent modifications of this technique, exemplified by the MarathonTM technology (Clontech Laboratories Inc.), for example, have significantly simplified the search for longer cDNAs. A slightly different technique, termed "restriction-site" PCR, uses universal primers to retrieve unknown nucleic acid sequence adjacent a known locus (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). Inverse PCR may also be used to amplify or to extend sequences using divergent primers based on a known region (Triglia, T., et al. (1988) Nucleic Acids Res. 16:8186). Another method which may be used is capture PCR which involves PCR amplification of DNA fragments adjacent a known sequence in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) PCR Methods Applic. 1: 111-119). Another method which may be used to retrieve unknown sequences is that of Parker, J.D. et al. (1991); Nucleic Acids Res. 19:3055-

3060). Additionally, one may use PCR, nested primers, and PromoterFinder™ libraries to walk genomic DNA (Clontech, Palo Alto, CA). This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

When screening for full-length cDNAs, it is preferable to use libraries that have been
5 size-selected to include larger cDNAs. Also, random-primed libraries are preferable, in that they will contain more sequences that contain the 5' regions of genes. Use of a randomly primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

10 In one embodiment of the invention, the nucleic acid molecules of the present invention may be used for chromosome localisation. In this technique, a nucleic acid molecule is specifically targeted to, and can hybridize with, a particular location on an individual human chromosome. The mapping of relevant sequences to chromosomes according to the present invention is an important step in the confirmatory correlation of those
15 sequences with the gene-associated disease. Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found in, for example, V. McKusick, Mendelian Inheritance in Man (available on-line through Johns Hopkins University Welch Medical Library). The relationships between genes and diseases that
20 have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes). This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localised by genetic linkage to a particular genomic region, any sequences mapping to that area may
25 represent associated or regulatory genes for further investigation. The nucleic acid molecule may also be used to detect differences in the chromosomal location due to translocation, inversion, etc. among normal, carrier, or affected individuals.

The nucleic acid molecules of the present invention are also valuable for tissue localisation. Such techniques allow the determination of expression patterns of the
30 polypeptide in tissues by detection of the mRNAs that encode them. These techniques

include in situ hybridization techniques and nucleotide amplification techniques, such as PCR. Results from these studies provide an indication of the normal functions of the polypeptide in the organism. In addition, comparative studies of the normal expression pattern of mRNAs with that of mRNAs encoded by a mutant gene provide valuable 5 insights into the role of mutant polypeptides in disease. Such inappropriate expression may be of a temporal, spatial or quantitative nature.

The vectors of the present invention comprise nucleic acid molecules of the invention and may be cloning or expression vectors. The host cells of the invention, which may be transformed, transfected or transduced with the vectors of the invention may be 10 prokaryotic or eukaryotic.

The polypeptides of the invention may be prepared in recombinant form by expression of their encoding nucleic acid molecules in vectors contained within a host cell. Such expression methods are well known to those of skill in the art and many are described in detail by Sambrook *et al* (supra) and Fernandez & Hoeffler (1998, eds. "Gene expression 15 systems. Using nature for the art of expression". Academic Press, San Diego, London, Boston, New York, Sydney, Tokyo, Toronto).

Generally, any system or vector that is suitable to maintain, propagate or express nucleic acid molecules to produce a polypeptide in the required host may be used. The appropriate nucleotide sequence may be inserted into an expression system by any of a 20 variety of well-known and routine techniques, such as, for example, those described in Sambrook *et al.*, (supra). Generally, the encoding gene can be placed under the control of a control element such as a promoter, ribosome binding site (for bacterial expression) and, optionally, an operator, so that the DNA sequence encoding the desired polypeptide is transcribed into RNA in the transformed host cell.

25 Examples of suitable expression systems include, for example, chromosomal, episomal and virus-derived systems, including, for example, vectors derived from: bacterial plasmids, bacteriophage, transposons, yeast episomes, insertion elements, yeast chromosomal elements, viruses such as baculoviruses, papova viruses such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, 30 or combinations thereof, such as those derived from plasmid and bacteriophage genetic

elements, including cosmids and phagemids. Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid.

Particularly suitable expression systems include microorganisms such as bacteria 5 transformed with recombinant bacteriophage, plasmid or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (for example, baculovirus); plant cell systems transformed with virus expression vectors (for example, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (for example, Ti or pBR322 10 plasmids); or animal cell systems. Cell-free translation systems can also be employed to produce the polypeptides of the invention.

Introduction of nucleic acid molecules encoding a polypeptide of the present invention into host cells can be effected by methods described in many standard laboratory 15 manuals, such as Davis et al., *Basic Methods in Molecular Biology* (1986) and Sambrook et al., *[supra]*. Particularly suitable methods include calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection (see Sambrook et al., 1989 *[supra]*; Ausubel et al., 1991 *[supra]*; Spector, 20 Goldman & Leinwald, 1998). In eukaryotic cells, expression systems may either be transient (for example, episomal) or permanent (chromosomal integration) according to the needs of the system.

The encoding nucleic acid molecule may or may not include a sequence encoding a control sequence, such as a signal peptide or leader sequence, as desired, for example, for secretion of the translated polypeptide into the lumen of the endoplasmic reticulum, into 25 the periplasmic space or into the extracellular environment. These signals may be endogenous to the polypeptide or they may be heterologous signals. Leader sequences can be removed by the bacterial host in post-translational processing.

In addition to control sequences, it may be desirable to add regulatory sequences that allow for regulation of the expression of the polypeptide relative to the growth of the host 30 cell. Examples of regulatory sequences are those which cause the expression of a gene to

be increased or decreased in response to a chemical or physical stimulus, including the presence of a regulatory compound or to various temperature or metabolic conditions. Regulatory sequences are those non-translated regions of the vector, such as enhancers, promoters and 5' and 3' untranslated regions. These interact with host cellular proteins to 5 carry out transcription and translation. Such regulatory sequences may vary in their strength and specificity. Depending on the vector system and host utilised, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the Bluescript phagemid (Stratagene, 10 LaJolla, CA) or pSport1TM plasmid (Gibco BRL) and the like may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (for example, heat shock, RUBISCO and storage 15 protein genes) or from plant viruses (for example, viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of the sequence, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

An expression vector is constructed so that the particular nucleic acid coding sequence is located in the vector with the appropriate regulatory sequences, the positioning and 20 orientation of the coding sequence with respect to the regulatory sequences being such that the coding sequence is transcribed under the "control" of the regulatory sequences, i.e., RNA polymerase which binds to the DNA molecule at the control sequences transcribes the coding sequence. In some cases it may be necessary to modify the 25 sequence so that it may be attached to the control sequences with the appropriate orientation; i.e., to maintain the reading frame.

The control sequences and other regulatory sequences may be ligated to the nucleic acid coding sequence prior to insertion into a vector. Alternatively, the coding sequence can be cloned directly into an expression vector that already contains the control sequences and an appropriate restriction site.

30 For long-term, high-yield production of a recombinant polypeptide, stable expression is

preferred. For example, cell lines which stably express the polypeptide of interest may be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells that successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

10 Mammalian cell lines available as hosts for expression are known in the art and include many immortalised cell lines available from the American Type Culture Collection (ATCC) including, but not limited to, Chinese hamster ovary (CHO), HeLa, baby hamster kidney (BHK), monkey kidney (COS), C127, 3T3, BHK, HEK 293, Bowes melanoma and human hepatocellular carcinoma (for example Hep G2) cells and a number of other cell lines.

In the baculovirus system, the materials for baculovirus/insect cell expression systems are commercially available in kit form from, inter alia, Invitrogen, San Diego CA (the "MaxBac" kit). These techniques are generally known to those skilled in the art and are described fully in Summers and Smith, Texas Agricultural Experiment Station Bulletin

20 No. 1555 (1987). Particularly suitable host cells for use in this system include insect cells such as *Drosophila S2* and *Spodoptera Sf9* cells.

There are many plant cell culture and whole plant genetic expression systems known in the art. Examples of suitable plant cellular genetic expression systems include those described in US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of 25 genetic expression in plant cell culture has been described by Zenk, (1991) *Phytochemistry* 30, 3861-3863.

In particular, all plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be utilised, so that whole plants are recovered which contain the transferred gene. Practically all plants can be regenerated from cultured cells or tissues, 30 including but not limited to all major species of sugar cane, sugar beet, cotton, fruit and

other trees, legumes and vegetables.

Examples of particularly preferred bacterial host cells include streptococci, staphylococci, *E. coli*, *Streptomyces* and *Bacillus subtilis* cells.

5 Examples of particularly suitable host cells for fungal expression include yeast cells (for example, *S. cerevisiae*) and *Aspergillus* cells.

Any number of selection systems are known in the art that may be used to recover transformed cell lines. Examples include the herpes simplex virus thymidine kinase (Wigler, M. *et al.* (1977) *Cell* 11:223-32) and adenine phosphoribosyltransferase (Lowy, I. *et al.* (1980) *Cell* 22:817-23) genes that can be employed in tk- or aprt \pm cells, 10 respectively.

Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dihydrofolate reductase (DHFR) that confers resistance to methotrexate (Wigler, M. *et al.* (1980) *Proc. Natl. Acad. Sci.* 77:3567-70); npt, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin, F. *et al* 15 (1981) *J. Mol. Biol.* 150:1-14) and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. Additional selectable genes have been described, examples of which will be clear to those of skill in the art.

Although the presence or absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For 20 example, if the relevant sequence is inserted within a marker gene sequence, transformed cells containing the appropriate sequences can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding a polypeptide of the invention under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates 25 expression of the tandem gene as well.

Alternatively, host cells that contain a nucleic acid sequence encoding a polypeptide of the invention and which express said polypeptide may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassays, for example,

fluorescence activated cell sorting (FACS) or immunoassay techniques (such as the enzyme-linked immunosorbent assay [ELISA] and radioimmunoassay [RIA]), that include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein (see Hampton, R. *et al.* (1990) Serological 5 Methods, a Laboratory Manual, APS Press, St Paul, MN) and Maddox, D.E. *et al.* (1983) *J. Exp. Med.*, 158, 1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labelled hybridization or PCR probes for detecting sequences related to nucleic acid 10 molecules encoding polypeptides of the present invention include oligolabelling, nick translation, end-labelling or PCR amplification using a labelled polynucleotide. Alternatively, the sequences encoding the polypeptide of the invention may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, 15 are commercially available, and may be used to synthesise RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3 or SP6 and labelled nucleotides. These procedures may be conducted using a variety of commercially available kits (Pharmacia & Upjohn, (Kalamazoo, MI); Promega (Madison WI); and U.S. Biochemical Corp., Cleveland, OH)).

Suitable reporter molecules or labels, which may be used for ease of detection, include 20 radionuclides, enzymes and fluorescent, chemiluminescent or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Nucleic acid molecules according to the present invention may also be used to create transgenic animals, particularly rodent animals. Such transgenic animals form a further aspect of the present invention. This may be done locally by modification of somatic 25 cells, or by germ line therapy to incorporate heritable modifications. Such transgenic animals may be particularly useful in the generation of animal models for drug molecules effective as modulators of the polypeptides of the present invention.

The polypeptide can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulphate or ethanol precipitation, acid extraction, 30 anion or cation exchange chromatography, phosphocellulose chromatography,

hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. High performance liquid chromatography is particularly useful for purification. Well known techniques for refolding proteins may be employed to regenerate an active conformation when the polypeptide is denatured during 5 isolation and or purification.

Specialised vector constructions may also be used to facilitate purification of proteins, as desired, by joining sequences encoding the polypeptides of the invention to a nucleotide sequence encoding a polypeptide domain that will facilitate purification of soluble proteins. Examples of such purification-facilitating domains include metal chelating 10 peptides such as histidine-tryptophan modules that allow purification on immobilised metals, protein A domains that allow purification on immobilised immunoglobulin, and the domain utilised in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, WA). The inclusion of cleavable linker sequences such as those specific for 15 Factor XA or enterokinase (Invitrogen, San Diego, CA) between the purification domain and the polypeptide of the invention may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing the polypeptide of the invention fused to several histidine residues preceding a thioredoxin or an 20 enterokinase cleavage site. The histidine residues facilitate purification by IMAC (immobilised metal ion affinity chromatography as described in Porath, J. *et al.* (1992) Prot. Exp. Purif. 3: 263-281) while the thioredoxin or enterokinase cleavage site provides a means for purifying the polypeptide from the fusion protein. A discussion of vectors 25 which contain fusion proteins is provided in Kroll, D.J. *et al.* (DNA Cell Biol. 199312:441-453).

If the polypeptide is to be expressed for use in screening assays, generally it is preferred 25 that it be produced at the surface of the host cell in which it is expressed. In this event, the host cells may be harvested prior to use in the screening assay, for example using techniques such as fluorescence activated cell sorting (FACS) or immunoaffinity techniques. If the polypeptide is secreted into the medium, the medium can be recovered 30 in order to recover and purify the expressed polypeptide. If polypeptide is produced intracellularly, the cells must first be lysed before the polypeptide is recovered.

The polypeptide of the invention can be used to screen libraries of compounds in any of a variety of drug screening techniques. Such compounds may activate (agonise) or inhibit (antagonise) the level of expression of the gene or the activity of the polypeptide of the invention and form a further aspect of the present invention. Preferred compounds are 5 effective to alter the expression of a natural gene which encodes a polypeptide of the first aspect of the invention or to regulate the activity of a polypeptide of the first aspect of the invention.

Agonist or antagonist compounds may be isolated from, for example, cells, cell-free preparations, chemical libraries or natural product mixtures. These agonists or antagonists 10 may be natural or modified substrates, ligands, enzymes, receptors or structural or functional mimetics. For a suitable review of such screening techniques, see Coligan et al., *Current Protocols in Immunology* 1(2):Chapter 5 (1991).

Compounds that are most likely to be good antagonists are molecules that bind to the polypeptide of the invention without inducing the biological effects of the polypeptide 15 upon binding to it. Potential antagonists include small organic molecules, peptides, polypeptides and antibodies that bind to the polypeptide of the invention and thereby inhibit or extinguish its activity. In this fashion, binding of the polypeptide to normal cellular binding molecules may be inhibited, such that the normal biological activity of the polypeptide is prevented.

20 The polypeptide of the invention that is employed in such a screening technique may be free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. In general, such screening procedures may involve using appropriate cells or cell membranes that express the polypeptide that are contacted with a test compound to observe binding, or stimulation or inhibition of a functional response. The functional 25 response of the cells contacted with the test compound is then compared with control cells that were not contacted with the test compound. Such an assay may assess whether the test compound results in a signal generated by activation of the polypeptide, using an appropriate detection system. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist in the presence of 30 the test compound is observed.

A preferred method for identifying an agonist or antagonist compound of a polypeptide of the present invention comprises:

- (a) contacting a cell expressing on the surface thereof the polypeptide according to the first aspect of the invention, the polypeptide being associated with a second component capable of providing a detectable signal in response to the binding of a compound to the polypeptide, with a compound to be screened under conditions to permit binding to the polypeptide; and
- 5 (b) determining whether the compound binds to and activates or inhibits the polypeptide by measuring the level of a signal generated from the interaction of the compound with the polypeptide.
- 10

A further preferred method for identifying an agonist or antagonist of a polypeptide of the invention comprises:

- (a) contacting a cell expressing on the surface thereof the polypeptide, the polypeptide being associated with a second component capable of providing a detectable signal in response to the binding of a compound to the polypeptide, with a compound to be screened under conditions to permit binding to the polypeptide; and
- 15 (b) determining whether the compound binds to and activates or inhibits the polypeptide by comparing the level of a signal generated from the interaction of the compound with the polypeptide with the level of a signal in the absence of the compound.
- 20 In further preferred embodiments, the general methods that are described above may further comprise conducting the identification of agonist or antagonist in the presence of labelled or unlabelled ligand for the polypeptide.

In another embodiment of the method for identifying an agonist or antagonist of a polypeptide of the present invention comprises:

- 25 determining the inhibition of binding of a ligand to cells which have a polypeptide of the invention on the surface thereof, or to cell membranes containing such a polypeptide, in the presence of a candidate compound under conditions to permit binding to the polypeptide, and determining the amount of ligand bound to the polypeptide. A compound capable of causing reduction of binding of a ligand is considered to be an

agonist or antagonist. Preferably the ligand is labelled.

More particularly, a method of screening for a polypeptide antagonist or agonist compound comprises the steps of:

- (a) incubating a labelled ligand with a whole cell expressing a polypeptide according to
5 the invention on the cell surface, or a cell membrane containing a polypeptide of the invention,
- (b) measuring the amount of labelled ligand bound to the whole cell or the cell membrane;
- (c) adding a candidate compound to a mixture of labelled ligand and the whole cell or the
10 cell membrane of step (a) and allowing the mixture to attain equilibrium;
- (d) measuring the amount of labelled ligand bound to the whole cell or the cell membrane
after step (c); and
- (e) comparing the difference in the labelled ligand bound in step (b) and (d), such that the
15 compound which causes the reduction in binding in step (d) is considered to be an agonist or antagonist.

In certain of the embodiments described above, simple binding assays may be used, in which the adherence of a test compound to a surface bearing the polypeptide is detected by means of a label directly or indirectly associated with the test compound or in an assay involving competition with a labelled competitor. In another embodiment, competitive
20 drug screening assays may be used, in which neutralising antibodies that are capable of binding the polypeptide specifically compete with a test compound for binding. In this manner, the antibodies can be used to detect the presence of any test compound that possesses specific binding affinity for the polypeptide.

Assays may also be designed to detect the effect of added test compounds on the
25 production of mRNA encoding the polypeptide in cells. For example, an ELISA may be constructed that measures secreted or cell-associated levels of polypeptide using monoclonal or polyclonal antibodies by standard methods known in the art, and this can be used to search for compounds that may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues. The formation of binding

complexes between the polypeptide and the compound being tested may then be measured.

Another technique for drug screening which may be used provides for high throughput screening of compounds having suitable binding affinity to the polypeptide of interest
5 (see International patent application WO84/03564). In this method, large numbers of different small test compounds are synthesised on a solid substrate, which may then be reacted with the polypeptide of the invention and washed. One way of immobilising the polypeptide is to use non-neutralising antibodies. Bound polypeptide may then be detected using methods that are well known in the art. Purified polypeptide can also be
10 coated directly onto plates for use in the aforementioned drug screening techniques.

The polypeptide of the invention may be used to identify membrane-bound or soluble receptors, through standard receptor binding techniques that are known in the art, such as ligand binding and crosslinking assays in which the polypeptide is labelled with a radioactive isotope, is chemically modified, or is fused to a peptide sequence that
15 facilitates its detection or purification, and incubated with a source of the putative receptor (for example, a composition of cells, cell membranes, cell supernatants, tissue extracts, or bodily fluids). The efficacy of binding may be measured using biophysical techniques such as surface plasmon resonance and spectroscopy. Binding assays may be used for the purification and cloning of the receptor, but may also identify agonists and
20 antagonists of the polypeptide, that compete with the binding of the polypeptide to its receptor. Standard methods for conducting screening assays are well understood in the art.

The invention also includes a screening kit useful in the methods for identifying agonists, antagonists, ligands, receptors, substrates, enzymes, that are described above.
25 The invention includes the agonists, antagonists, ligands, receptors, substrates and enzymes, and other compounds which modulate the activity or antigenicity of the polypeptide of the invention discovered by the methods that are described above.

The invention also provides pharmaceutical compositions comprising a polypeptide, nucleic acid, ligand or compound of the invention in combination with a suitable
30 pharmaceutical carrier. These compositions may be suitable as therapeutic or diagnostic

reagents, as vaccines, or as other immunogenic compositions, as outlined in detail below.

According to the terminology used herein, a composition containing a polypeptide, nucleic acid, ligand or compound [X] is "substantially free of" impurities [herein, Y] when at least 85% by weight of the total X+Y in the composition is X. Preferably, X 5 comprises at least about 90% by weight of the total of X+Y in the composition, more preferably at least about 95%, 98% or even 99% by weight.

The pharmaceutical compositions should preferably comprise a therapeutically effective amount of the polypeptide, nucleic acid molecule, ligand, or compound of the invention. The term "therapeutically effective amount" as used herein refers to an amount of a 10 therapeutic agent needed to treat, ameliorate, or prevent a targeted disease or condition, or to exhibit a detectable therapeutic or preventative effect. For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, for example, of neoplastic cells, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model may also be used to determine the appropriate concentration range and 15 route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

The precise effective amount for a human subject will depend upon the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and 20 tolerance/response to therapy. This amount can be determined by routine experimentation and is within the judgement of the clinician. Generally, an effective dose will be from 0.01 mg/kg to 50 mg/kg, preferably 0.05 mg/kg to 10 mg/kg. Compositions may be administered individually to a patient or may be administered in combination with other agents, drugs or hormones.

25 A pharmaceutical composition may also contain a pharmaceutically acceptable carrier, for administration of a therapeutic agent. Such carriers include antibodies and other polypeptides, genes and other therapeutic agents such as liposomes, provided that the carrier does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. 30 Suitable carriers may be large, slowly metabolised macromolecules such as proteins,

polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers and inactive virus particles.

5 Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulphates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable carriers is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

10 Pharmaceutically acceptable carriers in therapeutic compositions may additionally contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such compositions. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

15 Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated.

20 The pharmaceutical compositions utilised in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal or transcutaneous applications (for example, see WO98/20734), subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, intravaginal or rectal means. Gene guns or hyposprays may also be used to administer the pharmaceutical compositions of the invention. Typically, the therapeutic compositions may be prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to 25 injection may also be prepared.

Direct delivery of the compositions will generally be accomplished by injection, subcutaneously, intraperitoneally, intravenously or intramuscularly, or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Dosage treatment may be a single dose schedule or a multiple dose schedule.

If the activity of the polypeptide of the invention is in excess in a particular disease state, several approaches are available. One approach comprises administering to a subject an inhibitor compound (antagonist) as described above, along with a pharmaceutically acceptable carrier in an amount effective to inhibit the function of the polypeptide, such as by blocking the binding of ligands, substrates, enzymes, receptors, or by inhibiting a second signal, and thereby alleviating the abnormal condition. Preferably, such antagonists are antibodies. Most preferably, such antibodies are chimeric and/or humanised to minimise their immunogenicity, as described previously.

In another approach, soluble forms of the polypeptide that retain binding affinity for the ligand, substrate, enzyme, receptor, in question, may be administered. Typically, the polypeptide may be administered in the form of fragments that retain the relevant portions.

In an alternative approach, expression of the gene encoding the polypeptide can be inhibited using expression blocking techniques, such as the use of antisense nucleic acid molecules (as described above), either internally generated or separately administered. Modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5' or regulatory regions (signal sequence, promoters, enhancers and introns) of the gene encoding the polypeptide. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. *et al.* (1994) In: Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY). The complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes. Such oligonucleotides may be administered or may be generated *in situ* from expression *in vivo*.

In addition, expression of the polypeptide of the invention may be prevented by using ribozymes specific to its encoding mRNA sequence. Ribozymes are catalytically active RNAs that can be natural or synthetic (see for example Usman, N, *et al.*, *Curr. Opin.*

Struct. Biol (1996) 6(4), 527-33). Synthetic ribozymes can be designed to specifically cleave mRNAs at selected positions thereby preventing translation of the mRNAs into functional polypeptide. Ribozymes may be synthesised with a natural ribose phosphate backbone and natural bases, as normally found in RNA molecules. Alternatively the 5 ribozymes may be synthesised with non-natural backbones, for example, 2'-O-methyl RNA, to provide protection from ribonuclease degradation and may contain modified bases.

RNA molecules may be modified to increase intracellular stability and half-life. Possible 10 modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiester linkages within the backbone of the molecule. This concept is inherent 15 in the production of PNAs and can be extended in all of these molecules by the inclusion of non-traditional bases such as inosine, queosine and butosine, as well as acetyl-, methyl-, thio- and similarly modified forms of adenine, cytidine, guanine, thymine and uridine which are not as easily recognised by endogenous endonucleases.

For treating abnormal conditions related to an under-expression of the polypeptide of the invention and its activity, several approaches are also available. One approach comprises 20 administering to a subject a therapeutically effective amount of a compound that activates the polypeptide, i.e., an agonist as described above, to alleviate the abnormal condition. Alternatively, a therapeutic amount of the polypeptide in combination with a suitable pharmaceutical carrier may be administered to restore the relevant physiological balance 25 of polypeptide.

Gene therapy may be employed to effect the endogenous production of the polypeptide by the relevant cells in the subject. Gene therapy is used to treat permanently the 25 inappropriate production of the polypeptide by replacing a defective gene with a corrected therapeutic gene.

Gene therapy of the present invention can occur *in vivo* or *ex vivo*. *Ex vivo* gene therapy requires the isolation and purification of patient cells, the introduction of a therapeutic gene and introduction of the genetically altered cells back into the patient. In contrast, *in vivo* gene therapy does not require isolation and purification of a patient's cells.

The therapeutic gene is typically "packaged" for administration to a patient. Gene delivery vehicles may be non-viral, such as liposomes, or replication-deficient viruses, such as adenovirus as described by Berkner, K.L., in *Curr. Top. Microbiol. Immunol.*, 158, 39-66 (1992) or adeno-associated virus (AAV) vectors as described by Muzyczka,

5 N., in *Curr. Top. Microbiol. Immunol.*, 158, 97-129 (1992) and U.S. Patent No. 5,252,479. For example, a nucleic acid molecule encoding a polypeptide of the invention may be engineered for expression in a replication-defective retroviral vector. This expression construct may then be isolated and introduced into a packaging cell transduced with a retroviral plasmid vector containing RNA encoding the polypeptide, 10 such that the packaging cell now produces infectious viral particles containing the gene of interest. These producer cells may be administered to a subject for engineering cells in vivo and expression of the polypeptide in vivo (see Chapter 20, *Gene Therapy and other Molecular Genetic-based Therapeutic Approaches*, (and references cited therein) in *Human Molecular Genetics* (1996), T Strachan and A P Read, BIOS Scientific Publishers 15 Ltd).

Another approach is the administration of "naked DNA" in which the therapeutic gene is directly injected into the bloodstream or muscle tissue.

In situations in which the polypeptides or nucleic acid molecules of the invention are disease-causing agents, the invention provides that they can be used in vaccines to raise 20 antibodies against the disease causing agent.

Vaccines according to the invention may either be prophylactic (ie. to prevent infection) or therapeutic (ie. to treat disease after infection). Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with pharmaceutically-acceptable carriers as described above, which include 25 any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, and other pathogens.

30 Since polypeptides may be broken down in the stomach, vaccines comprising

polypeptides are preferably administered parenterally (for instance, subcutaneous, intramuscular, intravenous, or intradermal injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation 5 isotonic with the blood of the recipient, and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents.

The vaccine formulations of the invention may be presented in unit-dose or multi-dose containers. For example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. 10 The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

This invention also relates to the use of nucleic acid molecules according to the present invention as diagnostic reagents. Detection of a mutated form of the gene characterised by the nucleic acid molecules of the invention which is associated with a dysfunction will 15 provide a diagnostic tool that can add to, or define, a diagnosis of a disease, or susceptibility to a disease, which results from under-expression, over-expression or altered spatial or temporal expression of the gene. Individuals carrying mutations in the gene may be detected at the DNA level by a variety of techniques.

Nucleic acid molecules for diagnosis may be obtained from a subject's cells, such as from 20 blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR, ligase chain reaction (LCR), strand displacement amplification (SDA), or other amplification techniques (see Saiki *et al.*, *Nature*, 324, 163-166 (1986); Bej, *et al.*, *Crit. Rev. Biochem. Molec. Biol.*, 26, 301-334 (1991); Birkenmeyer *et al.*, *J. Virol. Meth.*, 35, 117-126 25 (1991); Van Brunt, *J. Bio/Technology*, 8, 291-294 (1990)) prior to analysis.

In one embodiment, this aspect of the invention provides a method of diagnosing a disease in a patient, comprising assessing the level of expression of a natural gene encoding a polypeptide according to the invention and comparing said level of expression to a control level, wherein a level that is different to said control level is indicative of 30 disease. The method may comprise the steps of:

a) contacting a sample of tissue from the patient with a nucleic acid probe under stringent conditions that allow the formation of a hybrid complex between a nucleic acid molecule of the invention and the probe;

b) contacting a control sample with said probe under the same conditions used in step a);

5 c) and detecting the presence of hybrid complexes in said samples;

wherein detection of levels of the hybrid complex in the patient sample that differ from levels of the hybrid complex in the control sample is indicative of disease.

A further aspect of the invention comprises a diagnostic method comprising the steps of:

a) obtaining a tissue sample from a patient being tested for disease;

10 b) isolating a nucleic acid molecule according to the invention from said tissue sample; and,

c) diagnosing the patient for disease by detecting the presence of a mutation in the nucleic acid molecule which is associated with disease.

To aid the detection of nucleic acid molecules in the above-described methods, an
15 amplification step, for example using PCR, may be included.

Deletions and insertions can be detected by a change in the size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to labelled RNA of the invention or alternatively, labelled antisense DNA sequences of the invention. Perfectly-matched sequences can be distinguished from

20 mismatched duplexes by RNase digestion or by assessing differences in melting temperatures. The presence or absence of the mutation in the patient may be detected by contacting DNA with a nucleic acid probe that hybridises to the DNA under stringent conditions to form a hybrid double-stranded molecule, the hybrid double-stranded molecule having an unhybridised portion of the nucleic acid probe strand at any portion
25 corresponding to a mutation associated with disease; and detecting the presence or absence of an unhybridised portion of the probe strand as an indication of the presence or absence of a disease-associated mutation in the corresponding portion of the DNA strand.

Such diagnostics are particularly useful for prenatal and even neonatal testing.

Point mutations and other sequence differences between the reference gene and "mutant" genes can be identified by other well-known techniques, such as direct DNA sequencing or single-strand conformational polymorphism, (see Orita *et al.*, *Genomics*, 5, 874-879 (1989)). For example, a sequencing primer may be used with double-stranded PCR 5 product or a single-stranded template molecule generated by a modified PCR. The sequence determination is performed by conventional procedures with radiolabelled nucleotides or by automatic sequencing procedures with fluorescent-tags. Cloned DNA segments may also be used as probes to detect specific DNA segments. The sensitivity of this method is greatly enhanced when combined with PCR. Further, point mutations and 10 other sequence variations, such as polymorphisms, can be detected as described above, for example, through the use of allele-specific oligonucleotides for PCR amplification of sequences that differ by single nucleotides.

DNA sequence differences may also be detected by alterations in the electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA 15 sequencing (for example, Myers *et al.*, *Science* (1985) 230:1242). Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (see Cotton *et al.*, *Proc. Natl. Acad. Sci. USA* (1985) 85: 4397-4401).

In addition to conventional gel electrophoresis and DNA sequencing, mutations such as 20 microdeletions, aneuploidies, translocations, inversions, can also be detected by *in situ* analysis (see, for example, Keller *et al.*, *DNA Probes*, 2nd Ed., Stockton Press, New York, N.Y., USA (1993)), that is, DNA or RNA sequences in cells can be analysed for mutations without need for their isolation and/or immobilisation onto a membrane. Fluorescence *in situ* hybridization (FISH) is presently the most commonly applied 25 method and numerous reviews of FISH have appeared (see, for example, Trachuck *et al.*, *Science*, 250: 559-562 (1990), and Trask *et al.*, *Trends, Genet.* 7:149-154 (1991)).

In another embodiment of the invention, an array of oligonucleotide probes comprising a nucleic acid molecule according to the invention can be constructed to conduct efficient screening of genetic variants, mutations and polymorphisms. Array technology methods 30 are well known and have general applicability and can be used to address a variety of

questions in molecular genetics including gene expression, genetic linkage, and genetic variability (see for example: M.Chee *et al.*, Science (1996) 274: 610-613).

In one embodiment, the array is prepared and used according to the methods described in PCT application WO95/11995 (Chee *et al.*; Lockhart, D. J. *et al.* (1996) Nat. Biotech. 14: 5 1675-1680); and Schena, M. *et al.* (1996) Proc. Natl. Acad. Sci. 93: 10614-10619). Oligonucleotide pairs may range from two to over one million. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support. In another aspect, an oligonucleotide may be synthesized 10 on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application WO95/251116 (Baldeschweiler *et al.*). In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, 15 such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8, 24, 96, 384, 1536 or 6144 oligonucleotides, or any other number between two and over one million which lends itself to the efficient use of commercially-available instrumentation.

20 In addition to the methods discussed above, diseases may be diagnosed by methods comprising determining, from a sample derived from a subject, an abnormally decreased or increased level of polypeptide or mRNA. Decreased or increased expression can be measured at the RNA level using any of the methods well known in the art for the quantitation of polynucleotides, such as, for example, nucleic acid amplification, for 25 instance PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods.

Assay techniques that can be used to determine levels of a polypeptide of the present invention in a sample derived from a host are well-known to those of skill in the art and are discussed in some detail above (including radioimmunoassays, competitive-binding 30 assays, Western Blot analysis and ELISA assays). This aspect of the invention provides a

diagnostic method which comprises the steps of: (a) contacting a ligand as described above with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and (b) detecting said complex.

Protocols such as ELISA, RIA, and FACS for measuring polypeptide levels may 5 additionally provide a basis for diagnosing altered or abnormal levels of polypeptide expression. Normal or standard values for polypeptide expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably humans, with antibody to the polypeptide under conditions suitable for complex formation. The amount of standard complex formation may be quantified by various 10 methods, such as by photometric means.

Antibodies which specifically bind to a polypeptide of the invention may be used for the diagnosis of conditions or diseases characterised by expression of the polypeptide, or in assays to monitor patients being treated with the polypeptides, nucleic acid molecules, ligands and other compounds of the invention. Antibodies useful for diagnostic purposes 15 may be prepared in the same manner as those described above for therapeutics. Diagnostic assays for the polypeptide include methods that utilise the antibody and a label to detect the polypeptide in human body fluids or extracts of cells or tissues. The antibodies may be used with or without modification, and may be labelled by joining them, either covalently or non-covalently, with a reporter molecule. A wide variety of 20 reporter molecules known in the art may be used, several of which are described above.

Quantities of polypeptide expressed in subject, control and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease. Diagnostic assays may be used to distinguish between absence, presence, and excess expression of polypeptide and to 25 monitor regulation of polypeptide levels during therapeutic intervention. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials or in monitoring the treatment of an individual patient.

A diagnostic kit of the present invention may comprise:

(a) a nucleic acid molecule of the present invention;

- (b) a polypeptide of the present invention; or
- (c) a ligand of the present invention.

In one aspect of the invention, a diagnostic kit may comprise a first container containing a nucleic acid probe that hybridises under stringent conditions with a nucleic acid molecule according to the invention; a second container containing primers useful for amplifying the nucleic acid molecule; and instructions for using the probe and primers for facilitating the diagnosis of disease. The kit may further comprise a third container holding an agent for digesting unhybridised RNA.

In an alternative aspect of the invention, a diagnostic kit may comprise an array of nucleic acid molecules, at least one of which may be a nucleic acid molecule according to the invention.

To detect polypeptide according to the invention, a diagnostic kit may comprise one or more antibodies that bind to a polypeptide according to the invention; and a reagent useful for the detection of a binding reaction between the antibody and the polypeptide.

Such kits will be of use in diagnosing a disease or susceptibility to diseases in which Nuclear Hormone Receptor Ligand Binding Domains are implicated, particularly cell proliferative disorders, including neoplasm, melanoma, lung, colorectal, breast, uterus, prostate, pancreas, head and neck and other solid tumours, myeloproliferative disorders, such as leukemia, non-Hodgkin lymphoma, leukopenia, thrombocytopenia, angiogenesis disorder, Kaposi's sarcoma, autoimmune/inflammatory disorders, including allergy, inflammatory bowel disease, arthritis, psoriasis and respiratory tract inflammation, asthma, and organ transplant rejection, cardiovascular disorders, including hypertension, hypotension, oedema, angina, atherosclerosis, thrombosis, sepsis, shock, reperfusion injury, heart arrhythmia, and ischemia, neurological disorders including, central nervous system disease, Alzheimer's disease, Parkinson's disease, brain injury, stroke, amyotrophic lateral sclerosis, anxiety, depression, and pain, cognition enhancement, learning and memory enhancement, developmental disorders, metabolic disorders including diabetes mellitus, osteoporosis, lipid metabolism disorder, hyperthyroidism, hyperparathyroidism, thyroid hormone resistance syndrome, hypercalcemia, hypocalcaemia, hypercholesterolemia, hyperlipidemia, and obesity, renal disorders,

including glomerulonephritis, renovascular hypertension, blood disorders including hemophilia, dermatological disorders, including, cellulite, acne, eczema, psoriasis and wound healing, scarring, negative effects of aging, fertility enhancement, contraception, pregnancy termination, progesterone antagonism, hormone replacement therapies, steroid 5 hormone-like mediated hair characteristics, immunomodulation, AIDS, vision disorders, glucocorticoid resistance, mineralocorticoid resistance, androgen resistance, pseudohypoaldosteronism, spinal/bulbar muscular atrophy, extraskeletal myxoid chondrosarcomas, adrenal insufficiency, sexual reversal, infections including viral 10 infection, bacterial infection, fungal infection and parasitic infection and other pathological conditions, particularly those in which nuclear hormone receptors are 15 implicated.

Various aspects and embodiments of the present invention will now be described in more detail by way of example, with particular reference to the LBDG5 polypeptide, the LBDG10 polypeptide, the LBDG11 polypeptide, the LBDG12 polypeptide, and the 15 LBDG13 polypeptide.

It will be appreciated that modification of detail may be made without departing from the scope of the invention.

Brief description of the Figures

Figure 1: Front page of the Biopendium™. Search initiated using 3ERT:A.
20 Figure 2A: Inpharmatica Genome Threader™ results of search using 3ERT:A. The arrow points to P06212, Chicken Estrogen Receptor a typical Nuclear Hormone Receptor Ligand Binding Domain family member.

Figure 2B: Inpharmatica Genome Threader™ results of search using 3ERT:A. The arrow points to the CAA05410.2 (LBDG5) protein.
25 Figure 2C: Inpharmatica PSI-Blast results from search using 3ERT:A. The arrow points to the CAA05410.2 (LBDG5) protein.

Figure 3: InterPro search results for CAA05410.2 (LBDG5).

Figure 4: NCBI Conserved Domain Database search results for CAA05410.2 (LBDG5).

Figure 5A: NCBI protein report for CAA05410.2 (LBDG5).

Figure 5B: SWISS-PROT protein report for O43245, an equivalent sequence to CAA05410.2 (LBDG5).

Figure 5C: Graphical view of NCBI PSI-BLAST iteration 1 to 10 results when queried with
5 CAA05410.2 (LBDG5).

Figure 5D: List of NCBI PSI-BLAST iteration 1 to 10 results when queried with
CAA05410.2 (LBDG5).

Figure 6: Inpharmatica Genome Threader™ results of search using CAA05410.2 (LBDG5) as the query sequence. The arrow points to 3ERT:A, the structure of Human Estrogen
10 Receptor alpha Ligand Binding Domain.

Figure 7A: Selection of Inpharmatica reverse-maximised PSI-Blast results from search using CAA05410.2 (LBDG5) as the query sequence. Relationships to known Nuclear Hormone Receptor Ligand Binding Domains are found in the second positive iteration. The arrow points to AAC52143.1, the Nuclear Hormone Receptor Ligand Binding Domain Rat
15 HZF-3.

Figure 7B: Selection of Inpharmatica PSI-Blast reverse-maximised results from search using CAA05410.2 (LBDG5) as the query sequence. The arrow points to 3ERT:A (Human Estrogen Receptor alpha Ligand Binding Domain), the original query structure.

Figure 8: Genome Threader™ alignment of CAA05410.2 (LBDG5) and 3ERT:A. Residues
20 which are conserved between CAA05410.2 (LBDG5) and 3ERT:A in the 3ERT:A dimerisation helix are boxed in black.

Figure 9: RasMol view of 3ERT:A (Human Estrogen Receptor Ligand Binding Domain) with residues in the dimerisation helix which are conserved between CAA05410.2 (LBDG5) and 3ERT:A highlighted in wireframe view of sidechain.

25 Figure 10: Inpharmatica PSI-BLAST results for CAA05410.2 (LBDG5), arrow 1 points to CAA05409.2 (LBDG10), the rat orthologue of CAA05410.2 (LBDG5). Arrow 2 points to BAB62888.1 (LBDG11), the human paralogue of CAA05410.2 (LBDG5). Arrow 3 points to AAH03486.1 (LBDG12), the mouse orthologue of BAB62888.1 (LBDG11). Arrow 4 points to AAK49953.1 (LBDG13), the rat orthologue of BAB62888.1 (LBDG11).

Figure 11: Inpharmatica Genome Threader™ results of search using CAA05409.2 (LBDG10) as the query sequence. The arrow points to 3ERT:A, the structure of Human Estrogen Receptor alpha Ligand Binding Domain.

5 Figure 12: Selection of Inpharmatica PSI-Blast results from search using CAA05409.2 (LBDG10) as the query sequence. The arrow points to 3ERT:A (Human Estrogen Receptor alpha Ligand Binding Domain).

Figure 13: Alignment of CAA05409.2 (LBDG10) with CAA05410.2 (LBDG5) and 3ERT:A. Residues which are conserved between CAA05409.2 (LBDG10), CAA05410.2 (LBDG5) and 3ERT:A in the 3ERT:A dimerisation helix are boxed in black.

10 Figure 14: InterPro search results for CAA05409.2 (LBDG10).

Figure 15: Conserved Domain (CD) Database search results for CAA05409.2 (LBDG10).

Figure 16A: Graphical view of NCBI PSI-BLAST iteration 1 to 10 results when queried with CAA05409.2 (LBDG10).

15 Figure 16B: List of NCBI PSI-BLAST iteration 1 to 10 results when queried with CAA05409.2 (LBDG10).

Figure 17: NCBI protein report for CAA05409.2 (LBDG10).

Figure 18: Diagram of relationships between 3ERT:A, CAA05410.2 (LBDG5), BAB62888.1 (LBDG11), AAH03486.1 (LBDG12) and AAK49953.1 (LBDG13).

20 Figure 19: Genome Threader alignment of 3ERT:A with CAA05410.2 (LBDG5). The sequences of BAB62888.1 (LBDG11), AAH03486.1 (LBDG12) and AAK49953.1 (LBDG13) have been added to the alignment with respect to CAA05410.2 (LBDG5). Grey boxes mark the boundaries of the Genome Threader relationship between 3ERT:A and CAA05410.2 (LBDG5).

Figure 20: InterPro search results for BAB62888.1 (LBDG11).

25 Figure 21: Conserved Domain (CD) Database search results for BAB62888.1 (LBDG11).

Figure 22: Graphical view of NCBI PSI-BLAST iteration 10 results when queried with BAB62888.1 (LBDG11).

Figure 23: List of NCBI PSI-BLAST iteration 10 results when queried with BAB62888.1 (LBDG11).

Figure 24: NCBI protein report for BAB62888.1 (LBDG11).

Figure 25: InterPro search results for AAH03486.1 (LBDG12).

5 Figure 26: Conserved Domain (CD) Database search results for AAH03486.1 (LBDG12).

Figure 27: Graphical view of NCBI PSI-BLAST iteration 10 results when queried with AAH03486.1 (LBDG12).

Figure 28: List of NCBI PSI-BLAST iteration 10 results when queried with AAH03486.1 (LBDG12).

10 Figure 29: NCBI protein report for AAH03486.1 (LBDG12).

Figure 30: InterPro search results for AAK49953.1 (LBDG13).

Figure 31: Conserved Domain (CD) Database search results for AAK49953.1 (LBDG13).

Figure 32: Graphical view of NCBI PSI-BLAST iteration 10 results when queried with AAK49953.1 (LBDG13).

15 Figure 33: List of NCBI PSI-BLAST iteration 10 results when queried with AAK49953.1 (LBDG13).

Figure 34: NCBI protein report for AAK49953.1 (LBDG13).

Examples

Example 1: CAA05410.2 (LBDG5)

In order to initiate a search for a novel, distantly related Nuclear Hormone Receptor Ligand Binding Domain, an archetypal Nuclear Hormone Receptor Ligand Binding Domain family member, Human Estrogen Receptor alpha Ligand Binding Domain is chosen. More specifically, the search is initiated using a structure from the Protein Data Bank (PDB) which is operated by the Research Collaboratory for Structural Bioinformatics.

The structure chosen is Human Estrogen Receptor alpha Ligand Binding Domain, PDB code 3ERT:A (Figure 1). A search of the Biopendium™ for homologues of 3ERT:A takes place and returns 6310 Inpharmatica Genome Threader™ results (selection given in Figure 2A and 2B) and 1093 Inpharmatica PSI-Blast results (selection in Figure 2C). The 6310 Genome Threader™ results include examples of other Nuclear Hormone Receptor Ligand Binding Domain family members, such as Chicken Estrogen Receptor, marked by an arrow in Figure 2A. Among the known Nuclear Hormone Receptor Ligand Binding Domain members appears a protein of apparently unknown function, CAA05410.2 (LBDG5, marked by an arrow in Figure 2B).

The Inpharmatica Genome Threader™ has identified residues 394-604 of a sequence, CAA05410.2 (LBDG5), as having an equivalent structure to residues 15-247 of Human Estrogen Receptor alpha Ligand Binding Domain (PDB code: 3ERT:A). Having a structure equivalent to 3ERT:A suggests that CAA05410.2 (LBDG5) is a protein that functions as a Nuclear Hormone Receptor Ligand Binding Domain. The Inpharmatica Genome Threader™ identifies this with 100% confidence.

Positive iterations of PSI-Blast are unable to identify this relationship (Forward PSI-Blast does identify known Nuclear Hormone Receptor Ligand Binding Domain members with varying degrees of probability (E-value) as would be expected.). It is only in negative iterations that Inpharmatica PSI-Blast can identify residues 441-518 of CAA05410.2 (LBDG5) as having a sequence relationship to the Human Estrogen Receptor alpha Ligand Binding Domain (Figure 2C). The ability to identify relationships via negative iterations of PSI-Blast is a product of the all-by-all sequence comparison (reverse-maximisation) that underlies the Biopendium and is unique to Inpharmatica. The identification of a relationship

between 3ERT:A and CAA05410.2 (LBDG5) in PSI-Blast iteration -5 at a significant E-value of 2.0E-08 strongly supports the Genome Threader annotation of CAA05410.2 (LBDG5) as containing a Nuclear Hormone Receptor Ligand Binding Domain.

In order to view the public domain annotation of CAA05410.2 (LBDG5) the InterPro secondary database is queried with CAA05410.2 (LBDG5; Figure 3). It can be seen from Figure 3 that InterPro annotates a region of CAA05410.2 (LBDG5) as containing a C4-type zinc finger. The region of CAA05410.2 (LBDG5) annotated by InterPro as containing a C4-type zinc finger lies between residues 303 and 363. This region is N-terminal to, and does not overlap with, residues 394-604 of CAA05410.2 (LBDG5), which Genome Threader and

Inpharmatica PSI-Blast annotate as containing a novel Nuclear Hormone Receptor Ligand Binding Domain. Thus residues 303-363 of CAA05410.2 (LBDG5) constitute a C4-type zinc finger. C4-type zinc fingers are often found N-terminal to Nuclear Hormone Receptor Ligand Binding Domains, and the observation of a C4-type zinc finger N-terminal to the proposed Nuclear Hormone Receptor Ligand Binding Domain further supports the Genome Threader annotation of CAA05410.2 (LBDG5) as containing a Nuclear Hormone Receptor Ligand Binding Domain. It is important to note, however, that the presence of a C4-type zinc finger does not indicate that a Nuclear Hormone Receptor Ligand Binding Domain must be present. Examples of well-documented proteins which contain a C4-type zinc finger but lack a Nuclear Hormone Receptor Ligand Binding Domain include Knirps (SWISS-PROT accession P10734) and ODR7 (SWISS-PROT accession P41933). Implicit in the existence of proteins such as Knirps and ODR7 is the fact that possession of a DBD does not mean that a LBD will be concomitantly present. It can be seen from Figure 3 that Interpro does not annotate any region of CAA05410.2 as containing a Nuclear Hormone Receptor Ligand Binding Domain. This demonstrates that CAA05410.2 (LBDG5) is unidentifiable as containing a Nuclear Hormone Receptor Ligand Binding Domain using InterPro.

In order to view what is known in the public domain secondary databases, the NCBI Conserved Domain Database (CDD) is queried with CAA05410.2 (LBDG5; Figure 4). CDD returns the C4-type zinc finger as discussed above but does not identify CAA05410.2 (LBDG5) as containing a Nuclear Hormone Receptor Ligand Binding Domain. This

demonstrates that CAA05410.2 (LBDG5) is unidentifiable as containing a Nuclear Hormone Receptor Ligand Binding Domain using CDD.

The National Centre for Biotechnology Information (NCBI) GenBank protein database is viewed to examine if there is any further information that is known in the public domain relating to CAA05410.2 (LBDG5). This is the U.S. public domain database for protein and gene sequence deposition (Figure 5A). CAA05410.2 was cloned by a group of scientists at the Anderson Cancer Centre, University of Texas, USA. The authors annotate CAA05410.2 (LBDG5) as "oncofetal protein p65" and classify it as a steroid/thyroid receptor superfamily member (a synonym for Nuclear Hormone Receptor family). However, this annotation does not reflect identification of a Ligand Binding Domain, instead this annotation of CAA05410.2 (LBDG5) as a Nuclear Hormone Receptor was performed solely on the basis of the "obvious" N-terminal C4-type zinc finger. No public domain annotation was made that CAA05410.2 (LBDG5) contains a Nuclear Hormone Receptor Ligand Binding Domain. Analogous classifications to the Nuclear Hormone receptor family have been made purely of the basis of a protein possessing a C4-type zinc finger (eg. Knirps and ODR7 are referred to as Nuclear Hormone Receptors). Implicit in the existence of proteins such as Knirps and ODR7 is the fact that possession of a C4-type zinc finger DBD does not mean that a LBD will be concomitantly present. Thus the (NCBI) GenBank does not in any way annotate CAA05410.2 (LBDG5) as containing a Nuclear Hormone Receptor Ligand Binding Domain or suggest that the protein has this function.

CAA05410.2 (LBDG5) is also deposited in the SWISS-PROT protein database under the identifier O43245 (Figure 5B). As discussed above, CAA05410.2 (LBDG5) is annotated as belonging to the Nuclear Hormone Receptors family, but there is no annotation that this sequence contains a Ligand Binding Domain. This is confirmed in the database cross-references section of the SWISS-PROT report (arrow Figure 5B) which annotates CAA05410.2 (LBDG5) as containing a C4-type zinc finger, but does not contain any reference or suggestion to the presence of a Nuclear Hormone Receptor Ligand Binding Domain.

NCBI provides a public domain PSI-Blast server. Querying NCBI PSI-Blast with CAA05410.2 (LBDG5) through 10 positive iterations fails to annotate any region of

CAA05410.2 (LBDG5) as having a relationship to any known Nuclear Hormone Receptor Ligand Binding Domains (note that (1) NCBI PSI-Blast has converged by iteration 10, and so further iterations would not return any more sequences and that (2) NCBI PSI-Blast cannot provide data on negative iterations because no all-by-all calculation is performed).

5 Figure 5C shows the graphical display of NCBI PSI-Blast results for CAA05410.2 (LBDG5). The horizontal axis corresponds to N-terminal to C-terminal residue numbering along the CAA05410.2 (LBDG5) protein. It is clear that there are only 6 sequences (lines Figure 5C) which are hit by the Nuclear Hormone Receptor Ligand Binding Domain region (residues 394-604) of CAA05410.2 (LBDG5). The accession codes of these 6 sequences in 10 are listed in Figure 5D (marked by lines). None of these 6 sequences have been annotated in the public domain as containing a Nuclear Hormone Receptor Ligand Binding Domain. Thus NCBI PSI-Blast does not annotate residues 394-604 of CAA05410.2 (LBDG5) as having a relationship to any known Nuclear Hormone receptor Ligand Binding Domains. Note that a large number of Nuclear Hormone Receptor sequences are hit by the region of' 15 CAA05410.2 (LBDG5) that contains an "obvious" C4-type zinc finger (in the region of residues 200-400 approximately). This indicates that NCBI PSI-Blast is annotating CAA05410.2 (LBDG5) as having a relationship to known Nuclear Hormone Receptor C4-type zinc fingers, but fails to find any relationships to any known Nuclear Hormone Receptor Ligand Binding Domains.

20 There is no further public domain annotation for CAA05410.2 (LBDG5) . The public domain information for this protein does not annotate it as containing a Nuclear Hormone Receptor Ligand Binding Domain. Therefore using all public domain annotation tools, CAA05410.2 (LBDG5) is not annotated as containing a Nuclear Hormone Receptor Ligand Binding Domain. Only the Inpharmatica Genome ThreaderTM and Inpharmatica PSI-Blast 25 are able to annotate this protein as containing a Nuclear Hormone Receptor Ligand Binding Domain.

The reverse search is now carried out. CAA05410.2 (LBDG5) is now used as the query sequence in the BiopendiumTM. The Inpharmatica Genome ThreaderTM identifies 157 hits (Figure 6) while Inpharmatica PSI-Blast returns 2093 hits (Figures 7A and 7B). The 30 Inpharmatica Genome ThreaderTM (Figure 6, arrow) identifies residues 394-604 of CAA05410.2 (LBDG5) as having a structure the same as Human Estrogen Receptor alpha

Ligand Binding Domain (PDB code: 3ERT:A) with 100% confidence. Thus a region from residues 394 to residues 604 of CAA05410.2 (LBDG5) has been identified as adopting an equivalent fold to the Human Estrogen Receptor alpha Ligand Binding Domain .

Inpharmatica PSI-Blast also identifies the same region of CAA05410.2 (LBDG5) as having 5 a relationship with known Nuclear Hormone receptor Ligand Binding Domains by the second positive iteration. For example, Figure 7A shows a selection of Inpharmatica PSI-Blast results and it can be seen that the sequence AAC52143.1 (Rat Nuclear Hormone Receptor HZF-3) has a highly significant relationship to CAA05410.2 (LBDG5), being found in the second positive iteration with an E-value of 1.0E-30. The results indicate that 10 residues 196-599 of CAA05410.2 (LBDG5) are related to residues 146-596 of AAC52143.1 (Rat Nuclear Hormone Receptor HZF-3). Residues 196-599 includes almost all of the CAA05410.2 (LBDG5) Nuclear Hormone Receptor Ligand Binding Domain region identified by Genome ThreaderTM (residues 394-604), and matches them to a region of HZF-3 which contains a known Nuclear Hormone Receptor Ligand Binding Domain 15 (residues 420-592, as determined by PFAM). Thus Inpharmatica PSI-Blast is in strong agreement with Inpharmatica Genome ThreaderTM at annotating a region between residues 394 to 604 of CAA05410.2 (LBDG5) as containing a Nuclear Hormone Receptor Ligand Binding Domain. This is in contrast to public domain NCBI PSI-Blast which fails to identify 20 any relationship between CAA05410.2 (LBDG5) and known Nuclear Hormone Receptor Ligand Binding Domains (Figures 5C and 5D). Only Inpharmatica Genome ThreaderTM and Inpharmatica PSI-Blast are able to identify CAA05410.2 (LBDG5) as containing a Nuclear Hormone Receptor Ligand Binding Domain. Inpharmatica PSI-Blast also identifies a 25 relationship between CAA05410.2 (LBDG5) and the original query structure 3ERT:A (Human Estrogen Receptor alpha Ligand Binding Domain), Figure 7B arrow. The relationship between CAA05410.2 (LBDG5) and 3ERT:A is found in the fifth positive iteration and has a significant E-value of 2.0E-08. This further consolidates the Genome Threader annotation of CAA05410.2 (LBDG5) as containing a Nuclear Hormone Receptor Ligand Binding Domain.

Among the Nuclear Hormone Receptor Ligand Binding Domain family members that the 30 Inpharmatica Genome ThreaderTM returns is the original input query Human Estrogen Receptor alpha Ligand Binding Domain (3ERT:A). 3ERT:A is chosen against which to

view the sequence alignment of CAA05410.2 (the LBDG5 polypeptide). Viewing the alignment (Figure 8) of the query protein against the protein identified as being of a similar structure helps to visualize the areas of homology. A particularly interesting feature of the Genome Threader alignment is found in the region that corresponds to the dimerisation helix of 3ERT:A (marked by asterisks in Figure 8). It can be seen that 12 residues of the 3ERT:A dimerisation helix are precisely conserved in CAA05410.2 (LBDG5; marked in black boxes). Particularly striking is the Ser-His-Ile-Arg-His-Met block of 100% identity. Because this helix is the primary interface for Ligand Binding Domain dimerisation, and plays a major role in determining dimer partner specificity, the conservation of residues in the dimerisation helix between 3ERT:A (Human Estrogen Receptor alpha LBD) and CAA05410.2 (LBDG5) strongly suggests that the Ligand Binding Domain of CAA05410.2 (LBDG5) will exhibit dimerisation properties similar to the Human Estrogen Receptor alpha Ligand Binding Domain. Since an Estrogen Receptor alpha Ligand Binding Domain can either homodimerise (ie. dimerise with another Estrogen Receptor alpha Ligand Binding Domain) or heterodimerise with an Estrogen Receptor beta Ligand Binding Domain, this would suggest that the Ligand Binding Domain of CAA05410.2 (LBDG5) will be able to (a) homodimerise with another CAA05410.2 (LBDG5) Ligand Binding Domain, (b) heterodimerise with an Estrogen Receptor alpha Ligand Binding Domain and (c) heterodimerise with an Estrogen Receptor beta Ligand Binding Domain. Predicted heterodimerisation with Estrogen Receptor alpha and Estrogen Receptor beta Ligand Binding Domains implicates CAA05410.2 (LBDG5) in all diseases linked to Estrogen Receptor alpha and Estrogen Receptor beta.

In order to ensure that the protein identified is a homologue of the query sequence, the visualisation program RasMol (Figure 9) is used. This visualization tools identifies important sites of known protein structures. This visualisation is shown with Human Estrogen Receptor alpha Ligand Binding Domain (3ERT:A), which illustrates the positions of residues on the dimerisation helix which are 100% conserved in CAA05410.2 (LBDG5).

To summarise, only Inpharmatica Genome ThreaderTM can identify that residues 394-604 of CAA05410.2 (LBDG5) folds in a similar manner to 3ERT:A (Human Estrogen Receptor alpha Ligand Binding Domain) and as such is identified as containing a novel Nuclear Hormone Receptor Ligand Binding Domain. This annotation is also strongly supported by

Inpharmatica PSI-Blast sequence-sequence relationships which relate this region of CAA05410.2 (LBDG5) to known Nuclear Hormone Receptor ligand Binding Domains.

Example 2: CAA05409.2 (LBDG10)

A *Rattus norvegicus* orthologue of CAA05410.2 (LBDG5) has been identified, and will be referred to as CAA05409.2 (LBDG10). CAA05409.2 (LBDG10) is identified in the first iteration of Inpharmatica PSI-BLAST (Figure 10, arrow 1) and has 83% sequence identity with CAA05410.2 (LBDG5). On the basis of the high homology to CAA05410.2 (LBDG5), we annotate CAA05409.2 (LBDG10) as also containing a Nuclear Hormone Receptor Ligand Binding Domain. In addition to this homology, CAA05409.2 (LBDG10) has Genome ThreaderTM and Inpharmatica PSI-BLAST relationships which consolidate the annotation of CAA05409.2 (LBDG10) as containing a Nuclear Hormone Receptor Ligand Binding Domain.

CAA05409.2 (LBDG10) is now used as the query sequence in the BiopendiumTM. The Inpharmatica Genome ThreaderTM identifies 128 hits (Figure 11) while Inpharmatica PSI-Blast returns 1920 hits (Figure 12). The Inpharmatica Genome ThreaderTM (Figure 11, arrow) identifies residues 413-603 of CAA05409.2 (LBDG10) as having a structure the same as Human Estrogen Receptor alpha Ligand Binding Domain (PDB code: 3ERT:A) with 100% confidence. Thus a region from residue 413 to residue 603 of CAA05409.2 (LBDG10) has been identified as adopting an equivalent fold to the Human Estrogen Receptor alpha Ligand Binding Domain.

Inpharmatica PSI-Blast also identifies the same region of CAA05409.2 (LBDG10) as having a relationship with known Nuclear Hormone receptor Ligand Binding Domains by the fourth positive iteration. For example, Figure 12 shows a selection of Inpharmatica PSI-Blast results and it can be seen that the sequence of 3ERT:A (Human Estrogen Receptor alpha Ligand Binding Domain) has a highly significant relationship to CAA05409.2 (LBDG10), being found in the fourth positive iteration with an E-value of 2.0E⁻⁴³. The results indicate that residues 413-603 of CAA05409.2 (LBDG10) are related to 3ERT:A (Human Estrogen Receptor alpha Ligand Binding Domain). Thus Inpharmatica PSI-Blast is in complete agreement with Inpharmatica Genome ThreaderTM at annotating a region

between residues 413 to 603 of CAA05409.2 (LBDG10) as containing a Nuclear Hormone Receptor Ligand Binding Domain.

Among the Nuclear Hormone Receptor Ligand Binding Domain family members that the Inpharmatica Genome Threader™ returns is the Human Estrogen Receptor alpha Ligand Binding Domain (3ERT:A). 3ERT:A is chosen against which to view the sequence alignment of CAA05409.2 (the LBDG10 polypeptide). CAA05410.2 (LBDG5) is also included in the alignment. Viewing the alignment (Figure 13) of the query protein against the protein identified as being of a similar structure helps to visualize the areas of homology. A particularly interesting feature of the Genome Threader alignment is found in the region that corresponds to the dimerisation helix of 3ERT:A (marked by asterisks in Figure 13). It can be seen that 12 residues of the 3ERT:A dimerisation helix are precisely conserved in CAA05409.2 (LBDG10; marked in black boxes). Particularly striking is the Ser-His-Ile-Arg-His-Met block of 100% identity. Because this helix is the primary interface for Ligand Binding Domain dimerisation, and plays a major role in determining dimer partner specificity, the conservation of residues in the dimerisation helix between 3ERT:A (Human Estrogen Receptor alpha LBD) and CAA05409.2 (LBDG10) strongly suggests that the Ligand Binding Domain of CAA05409.2 (LBDG10) will exhibit dimerisation properties similar to the Human Estrogen Receptor alpha Ligand Binding Domain. Since an Estrogen Receptor alpha Ligand Binding Domain can either homodimerise (ie. dimerise with another Estrogen Receptor alpha Ligand Binding Domain) or heterodimerise with an Estrogen Receptor beta Ligand Binding Domain, this would suggest that in *Rattus norvegicus* the Ligand Binding Domain of CAA05409.2 (LBDG10) will be able to (a) homodimerise with another CAA05409.2 (LBDG10) Ligand Binding Domain, (b) heterodimerise with a (Rat) Estrogen Receptor alpha Ligand Binding Domain and (c) heterodimerise with a (Rat) Estrogen Receptor beta Ligand Binding Domain. This also suggests that in *Homo sapiens* cell-based experimental systems the Ligand Binding Domain of CAA05409.2 (LBDG10) will be able to (d) heterodimerise with the Human orthologue CAA05410.2 (LBDG5) Ligand Binding Domain, (e) heterodimerise with a (Human) Estrogen Receptor alpha Ligand Binding Domain and (f) heterodimerise with a (Human) Estrogen Receptor beta Ligand Binding Domain. Predicted heterodimerisation with Estrogen Receptor alpha and

Estrogen Receptor beta Ligand Binding Domains implicates CAA05409.2 (LBDG10) in all diseases linked to Estrogen Receptor alpha and Estrogen Receptor beta.

In order to view the public domain annotation of CAA05409.2 (LBDG10) the InterPro secondary database is queried with CAA05409.2 (LBDG10; Figure 14). It can be seen from 5 Figure 14 that InterPro annotates a region of CAA05409.2 (LBDG10) as containing a C4-type zinc finger. The region of CAA05409.2 (LBDG10) annotated by InterPro as containing a C4-type zinc finger lies between residues 301 and 369. This region is N-terminal to, and does not overlap with, residues 413-603 of CAA05409.2 (LBDG10), which Genome Threader and Inpharmatica PSI-Blast annotate as containing a novel Nuclear Hormone 10 Receptor Ligand Binding Domain. Thus residues 301-369 of CAA05409.2 (LBDG10) constitute a C4-type zinc finger. C4-type zinc fingers are often found N-terminal to Nuclear Hormone Receptor Ligand Binding Domains, and the observation of a C4-type zinc finger N-terminal to the proposed Nuclear Hormone Receptor Ligand Binding Domain further supports the Genome Threader annotation of of CAA05409.2 (LBDG10) as containing a 15 Nuclear Hormone Receptor Ligand Binding Domain. It is important to note, however, that the presence of a C4-type zinc finger does not indicate that a Nuclear Hormone Receptor Ligand Binding Domain must be present. Examples of well-documented proteins which contain a C4-type zinc finger but lack a Nuclear Hormone Receptor Ligand Binding Domain include Knirps (SWISS-PROT accession P10734) and ODR7 (SWISS-PROT accession P41933). Implicit in the existence of proteins such as Knirps and ODR7 is the fact 20 that possession of a DBD does not mean that a LBD will be concomitantly present. It can be seen from Figure 14 that Interpro does not anotate any region of CAA05409.2 (LBDG10) as containing a Nuclear Hormone Receptor Ligand Binding Domain. This demonstrates that 25 CAA05409.2 (LBDG10) is unidentifiable as containing a Nuclear Hormone Receptor Ligand Binding Domain using InterPro.

In order to view what is known in the public domain secondary databases, the NCBI Conserved Domain Database (CDD) is queried with CAA05409.2 (LBDG10; Figure 15). CDD returns the C4-type zinc finger as discussed above but does not identify CAA05409.2 (LBDG10) as containing a Nuclear Hormone Receptor Ligand Binding Domain. This

demonstrates that CAA05409.2 (LBDG10) is unidentifiable as containing a Nuclear Hormone Receptor Ligand Binding Domain using CDD.

NCBI provides a public domain PSI-Blast server. Querying NCBI PSI-Blast with CAA05409.2 (LBDG10) through 10 positive iterations fails to annotate any region of CAA05409.2 (LBDG10) as having a relationship to any known Nuclear Hormone Receptor Ligand Binding Domains (note that NCBI PSI-Blast cannot provide data on negative iterations because no all-by-all calculation is performed). Figure 16A shows the graphical display of NCBI PSI-Blast results for CAA05409.2 (LBDG10). The horizontal axis corresponds to N-terminal to C-terminal residue numbering along the CAA05409.2 (LBDG10) protein. It is clear that there are only 6 sequences (lines Figure 16A) which are hit by the Nuclear Hormone Receptor Ligand Binding Domain region (residues 413-603) of CAA05409.2 (LBDG10). The accession codes of these 6 sequences in are listed in Figure 16B (marked by lines). None of these 6 sequences have been annotated in the public domain as containing a Nuclear Hormone Receptor Ligand Binding Domain. Thus NCBI PSI-Blast does not annotate residues 413-603 of CAA05409.2 (LBDG10) as having a relationship to any known Nuclear Hormone receptor Ligand Binding Domains. Note that a large number of Nuclear Hormone Receptor sequences are hit by the region of CAA05409.2 (LBDG10) that contains an "obvious" C4-type zinc finger (in the region of residues 200-400 approximately). This indicates that NCBI PSI-Blast is annotating CAA05409.2 (LBDG10) as having a relationship to known Nuclear Hormone Receptor C4-type zinc fingers, but fails to find any relationships to any known Nuclear Hormone Receptor Ligand Binding Domains.

The National Centre for Biotechnology Information (NCBI) GenBank protein database is viewed to examine if there is any further information that is known in the public domain relating to CAA05409.2 (LBDG10). This is the U.S. public domain database for protein and gene sequence deposition (Figure 17). CAA05409.1 was cloned by a group of scientists at the Anderson Cancer Centre, University of Texas, USA. The authors annotate CAA05409.2 (LBDG10) as "oncofetal protein p65" and classify it as a steroid/thyroid receptor superfamily member (a synonym for Nuclear Hormone Receptor family). However, this annotation does not reflect identification of a Ligand Binding

Domain, instead this annotation of CAA05409.2 (LBDG10) as a Nuclear Hormone Receptor was performed solely on the basis of the "obvious" N-terminal C4-type zinc finger. No public domain annotation was made that CAA05409.2 (LBDG10) contains a Nuclear Hormone Receptor Ligand Binding Domain. Analogous classifications to the 5 Nuclear Hormone receptor family have been made purely of the basis of a protein possessing a C4-type zinc finger (eg. Knirps and ODR7 are referred to as Nuclear Hormone Receptors). Implicit in the existence of proteins such as Knirps and ODR7 is the fact that possession of a C4-type zinc finger DBD does not mean that a LBD will be concomitantly present. Thus the (NCBI) GenBank does not in any way annotate 10 CAA05409.2 (LBDG10) as containing a Nuclear Hormone Receptor Ligand Binding Domain or suggest that the protein has this function.

There is no further public domain annotation for CAA05409.2 (LBDG10). The public domain information for this protein does not annotate it as containing a Nuclear Hormone Receptor Ligand Binding Domain. Therefore using all public domain annotation tools, 15 CAA05409.2 (LBDG10) is not annotated as containing a Nuclear Hormone Receptor Ligand Binding Domain.

To summarise, Inpharmatica Genome ThreaderTM can identify that residues 413-603 of CAA05409.2 (LBDG10) folds in a similar manner to 3ERT:A (Human Estrogen Receptor alpha Ligand Binding Domain) and as such is identified as containing a novel Nuclear 20 Hormone Receptor Ligand Binding Domain. This annotation is also strongly supported by Inpharmatica PSI-Blast sequence-sequence relationships which relate this region of CAA05409.2 (LBDG10) to known Nuclear Hormone Receptor ligand Binding Domains. This is further consolidated by the observation that CAA05409.2 (LBDG10) has high sequence homology to CAA05410.2 (LBDG5).

25 **Example 3: BAB62888.1 (LBDG11)**

A *Homö sapiens* parologue of CAA05410.2 (LBDG5) has been identified, and will be referred to herein as BAB62888.1 (LBDG11). Residues 52-329 of BAB62888.1 (LBDG11) are identified in the first iteration of Inpharmatica PSI-BLAST (Figure 10, arrow 2) as sharing 60% sequence identity with residues 330-614 of CAA05410.2 (LBDG5). Residues

330-614 of CAA05410.2 (LBDG5) contain the region (residues 394-604) predicted to adopt the structure of a Nuclear Hormone Receptor Ligand Binding Domain. On the basis of the high sequence identity that BAB62888.1 (LBDG11) shares with the region of CAA05410.2 (LBDG5) that is predicted to adopt the structure of a Nuclear Hormone Receptor Ligand

5 Binding Domain, we predict that BAB62888.1 (LBDG11) also adopts the structure of a Nuclear Hormone Receptor Ligand Binding Domain. Chothia and Lesk, 1986 (EMBO Journal vol.5 pp823) first showed that for proteins with more than 50% sequence identity 85% of residues would adopt the same conformation. Other groups (Sander, C. and Schneider, R. (1991) Proteins vol.9 pp56; Hubbard, T.J.P. and Blundell, T.L. (1987) Protein

10 Engineering vol.1 pp159; Flores, T.P., Orengo, C.A., Moss, D.M. and Thornton, J.M. (1993) Protein Science vol.2 pp1811; and Hilbert, M., Bohm, G. and Jaenicke, R. (1993) Proteins vol.17 pp138) subsequently extended these studies and have showed that the fold remains the same even if sequence identity falls as low as 30%.

Figure 18 schematically depicts the relationships between the structure of the Human
15 Estrogen Receptor alpha Ligand Binding Domain (3ERT:A), CAA05410.2 (LBDG5) and
BAB62888.1 (LBDG11). Genome Threader™ and Inpharmatica PSI-BLAST have
identified residues 394-604 of CAA05410.2 (LBDG5) as adopting the structure of residues
15-247 (residues 320-552 in full-length numbering scheme) of the Human Estrogen
Receptor alpha Ligand Binding Domain (3ERT:A), double-headed arrow at the top of
20 Figure 18. The relationships between 3ERT:A, CAA05410.2 (LBDG5) and BAB62888.1
(LBDG11) are also presented in the form of a multiple alignment (Figure19). This
alignment has been generated by first taking the Genome Threader alignment of 3ERT:A
with CAA05410.2 (LBDG5), and then adding the sequence of BAB62888.1 (LBDG11) to
the alignment on the basis of it's sequence identity to CAA05410.2 (LBDG5). By reference
25 to the alignment it is possible to map the domain boundaries of the predicted Ligand
Binding Domain of CAA05410.2 (LBDG5) onto BAB62888.1 (LBDG11): these domain
boundary residues are marked by grey boxes in Figure 19. This identifies that residues 118-
319 of BAB62888.1 (LBDG11) correspond to the residues 394-604 of CAA05410.2
(LBDG5) that are predicted to adopt the structure of a Ligand Binding Domain by Genome
30 Threader and Inpharmatica PSI-BLAST (see also Figure 18). Thus we annotate residues
118-319 of BAB62888.1 (LBDG11) as adopting the structure of a Nuclear Hormone

Receptor Ligand Binding Domain on the basis of sharing high sequence homology with residues 394-604 of CAA05410.2 (LBDG5).

In order to view the public domain annotation of BAB62888.1 (LBDG11) the InterPro secondary database is queried with BAB62888.1 (LBDG11; Figure 20). It can be seen from 5 Figure 20 that no matches are found to InterPro. Returning no matches demonstrates that BAB62888.1 (LBDG11) is unidentifiable as containing a Nuclear Hormone Receptor Ligand Binding Domain using InterPro.

In order to view what is known in the public domain secondary databases, the NCBI Conserved Domain Database (CDD) is queried with BAB62888.1 (LBDG11; Figure 21). 10 CDD returns no hits. Returning no hits demonstrates that BAB62888.1 (LBDG11) is unidentifiable as containing a Nuclear Hormone Receptor Ligand Binding Domain using CDD.

NCBI provides a public domain PSI-Blast server. Querying NCBI PSI-Blast with BAB62888.1 (LBDG11) through 10 positive iterations fails to annotate any region of 15 BAB62888.1 (LBDG11) as having a relationship to any known Nuclear Hormone Receptor Ligand Binding Domains (note that NCBI PSI-Blast cannot provide data on negative iterations because no all-by-all calculation is performed). Figure 22 shows the graphical display of NCBI PSI-Blast results for BAB62888.1 (LBDG11). Figure 23 lists these NCBI PSI-BLAST hits (marked by lines). None of these sequences have been annotated in the 20 public domain as containing a Nuclear Hormone Receptor Ligand Binding Domain. Thus NCBI PSI-Blast does not annotate BAB62888.1 (LBDG11) as having a relationship to any known Nuclear Hormone receptor Ligand Binding Domains.

The National Centre for Biotechnology Information (NCBI) GenBank protein database is viewed to examine if there is any further information that is known in the public domain 25 relating to BAB62888.1 (LBDG11). This is the U.S. public domain database for protein and gene sequence deposition (Figure 24). BAB62888.1 was cloned by a group of scientists at the Science University of Tokyo, Japan. The authors annotate BAB62888.1 (LBDG11) as a protein that interacts with Terminal Deoxynucleotidyl Transferase (TdT). TdT is a DNA polymerase that enhances Ig and TcR gene diversity in the N region in B- and T-cells. 30 Annotation as a TdT interacting protein has no connection to the Nuclear Hormone

Receptor Ligand Binding Domain superfamily. However, informed by the Inpharmatica annotation of BAB62888.1 (LBDG11) as containing a novel Nuclear Hormone Receptor Ligand Binding Domain we are able to predict that BAB62888.1 (LBDG11) will regulate TdT activity in response to steroid-like ligands. Thus the (NCBI) GenBank does not in any 5 way annotate BAB62888.1 (LBDG11) as containing a Nuclear Hormone Receptor Ligand Binding Domain or suggest that the protein has this function.

There is no further public domain annotation for BAB62888.1 (LBDG11) . The public domain information for this protein does not annotate it as containing a Nuclear Hormone Receptor Ligand Binding Domain. Therefore using all public domain annotation tools, 10 BAB62888.1 (LBDG11) is not annotated as containing a Nuclear Hormone Receptor Ligand Binding Domain. To summarise we annotate residues 118-319 of BAB62888.1 (LBDG11) as adopting the structure of a Nuclear Hormone Receptor Ligand Binding Domain on the basis of sharing high sequence homology with residues 394-604 of CAA05410.2 (LBDG5) which Genome Threader and Inpharmatica PSI-BLAST predict will 15 adopt the structure of a Nuclear Hormone Receptor Ligand Binding Domain.

Example 4: AAH03486.1 (LBDG12)

A *Mus musculus* orthologue of BAB62888.1 (LBDG11) has been identified and will be referred to herein as AAH03486.1 (LBDG12). Residues 52-328 of AAH03486.1 (LBDG12) are identified in the first iteration of Inpharmatica PSI-BLAST (Figure 10, arrow 3) as 20 sharing 62% sequence identity with residues 330-614 of CAA05410.2 (LBDG5). Residues 330-614 of CAA05410.2 (LBDG5) contain the region (residues 394-604) predicted to adopt the structure of a Nuclear Hormone Receptor Ligand Binding Domain. On the basis of the high sequence identity that AAH03486.1 (LBDG12) shares with the region of CAA05410.2 (LBDG5) that is predicted to adopt the structure of a Nuclear Hormone Receptor Ligand 25 Binding Domain, we predict that AAH03486.1 (LBDG12) also adopts the structure of a Nuclear Hormone Receptor Ligand Binding Domain. Chothia and Lesk, 1986 (EMBO Journal vol.5 pp823) first showed that for proteins with more than 50% sequence identity 85% of residues would adopt the same conformation. Other groups (Sander, C. and Schneider, R. (1991) Proteins vol.9 pp56; Hubbard, T.J.P. and Blundell, T.L. (1987) Protein 30 Engineering vol.1 pp159; Flores, T.P., Orengo, C.A., Moss, D.M. and Thornton, J.M.

(1993) Protein Science vol.2 pp1811; and Hilbert, M., Bohm, G. and Jaenicke, R. (1993) Proteins vol.17 pp138) subsequently extended these studies and have showed that the fold remains the same even if sequence identity falls as low as 30%.

Figure 18 schematically depicts the relationships between the structure of the Human Estrogen Receptor alpha Ligand Binding Domain (3ERT:A), CAA05410.2 (LBDG5) and AAH03486.1 (LBDG12). Genome Threader™ and Inpharmatica PSI-BLAST have identified residues 394-604 of CAA05410.2 (LBDG5) as adopting the structure of residues 15-247 of the Human Estrogen Receptor alpha Ligand Binding Domain (3ERT:A), double-headed arrow at the top of Figure 18. The relationships between (3ERT:A), CAA05410.2 (LBDG5) and AAH03486.1 (LBDG12) are also presented in the form of a multiple alignment (Figure19). This alignment has been generated by first taking the Genome Threader alignment of 3ERT:A with CAA05410.2 (LBDG5), and then adding the sequence of AAH03486.1 (LBDG12) to the alignment on the basis of it's sequence identity to CAA05410.2 (LBDG5). By reference to the alignment it is possible to map the domain boundaries of the predicted Ligand Binding Domain of CAA05410.2 (LBDG5) onto AAH03486.1 (LBDG12), these domain boundary residues are marked by grey boxes in Figure 19. This identifies that residues 118-318 of AAH03486.1 (LBDG12) correspond to the residues 394-604 of CAA05410.2 (LBDG5) that are predicted to adopt the structure of a Ligand Binding Domain by Genome Threader and Inpharmatica PSI-BLAST (see also Figure 18). Thus we annotate residues 118-318 of AAH03486.1 (LBDG12) as adopting the structure of a Nuclear Hormone Receptor Ligand Binding Domain on the basis of sharing high sequence homology with residues 394-604 of CAA05410.2 (LBDG5).

In order to view the public domain annotation of AAH03486.1 (LBDG12) the InterPro secondary database is queried with AAH03486.1 (LBDG12; Figure 25). It can be seen from Figure 25 that no matches are found to InterPro. Returning no matches demonstrates that AAH03486.1 (LBDG12) is unidentifiable as containing a Nuclear Hormone Receptor Ligand Binding Domain using InterPro.

In order to view what is known in the public domain secondary databases, the NCBI Conserved Domain Database (CDD) is queried with AAH03486.1 (LBDG12; Figure 26). CDD returns no hits. Returning no hits demonstrates that AAH03486.1 (LBDG12) is

unidentifiable as containing a Nuclear Hormone Receptor Ligand Binding Domain using CDD.

NCBI provides a public domain PSI-Blast server. Querying NCBI PSI-Blast with AAH03486.1 (LBDG12) through 10 positive iterations fails to annotate any region of 5 AAH03486.1 (LBDG12) as having a relationship to any known Nuclear Hormone Receptor Ligand Binding Domains (note that NCBI PSI-Blast cannot provide data on negative iterations because no all-by-all calculation is performed). Figure 27 shows the graphical display of NCBI PSI-Blast results for AAH03486.1 (LBDG12). Figure 28 lists these NCBI PSI-BLAST hits (marked by lines). None of these sequences have been annotated in the 10 public domain as containing a Nuclear Hormone Receptor Ligand Binding Domain. Thus NCBI PSI-Blast does not annotate AAH03486.1 (LBDG12) as having a relationship to any known Nuclear Hormone receptor Ligand Binding Domains.

The National Centre for Biotechnology Information (NCBI) GenBank protein database is viewed to examine if there is any further information that is known in the public domain 15 relating to AAH03486.1 (LBDG12). This is the U.S. public domain database for protein and gene sequence deposition (Figure 29). AAH03486.1 was cloned by the I.M.A.G.E. Consortium. AAH03486.1 is annotated as being similar to p65 protein, but as discussed earlier, p65 (CAA05410.2 (LBDG5)) is not annotated in the public domain as containing a Nuclear Hormone Receptor Ligand Binding Domain. Thus the (NCBI) GenBank does not in 20 any way annotate AAH03486.1 (LBDG12) as containing a Nuclear Hormone Receptor Ligand Binding Domain or suggest that the protein has this function.

There is no further public domain annotation for AAH03486.1 (LBDG12) . The public domain information for this protein does not annotate it as containing a Nuclear Hormone Receptor Ligand Binding Domain. Therefore using all public domain annotation tools, 25 AAH03486.1 (LBDG12) is not annotated as containing a Nuclear Hormone Receptor Ligand Binding Domain.

To summarise we annotate residues 118-318 of AAH03486.1 (LBDG12) as adopting the structure of a Nuclear Hormone Receptor Ligand Binding Domain on the basis of sharing high sequence homology with residues 394-604 of CAA05410.2 (LBDG5) which Genome

Threader and Inpharmatica PSI-BLAST predict will adopt the structure of a Nuclear Hormone Receptor Ligand Binding Domain.

Example 5: AAK49953.1 (LBDG13)

A *Rattus norvegicus* orthologue of BAB62888.1 (LBDG11) has been identified and will be referred to herein as AAK49953.1 (LBDG13). Residues 51-327 of AAK49953.1 (LBDG13) are identified in the first iteration of Inpharmatica PSI-BLAST (Figure 10, arrow 4) as sharing 59% sequence identity with residues 330-614 of CAA05410.2 (LBDG5). Residues 330-614 of CAA05410.2 (LBDG5) contain the region (residues 394-604) predicted to adopt the structure of a Nuclear Hormone Receptor Ligand Binding Domain. On the basis of the high sequence identity that AAK49953.1 (LBDG13) shares with the region of CAA05410.2 (LBDG5) that is predicted to adopt the structure of a Nuclear Hormone Receptor Ligand Binding Domain, we predict that AAK49953.1 (LBDG13) also adopts the structure of a Nuclear Hormone Receptor Ligand Binding Domain. Chothia and Lesk, 1986 (EMBO Journal vol.5 pp823) first showed that for proteins with more than 50% sequence identity 85% of residues would adopt the same conformation. Other groups (Sander, C. and Schneider, R. (1991) Proteins vol.9 pp56; Hubbard, T.J.P. and Blundell, T.L. (1987) Protein Engineering vol.1 pp159; Flores, T.P., Orengo, C.A., Moss, D.M. and Thornton, J.M. (1993) Protein Science vol.2 pp1811; and Hilbert, M., Bohm, G. and Jaenicke, R. (1993) Proteins vol.17 pp138) subsequently extended these studies and have showed that the fold remains the same even if sequence identity falls as low as 30%.

Figure 18 schematically depicts the relationships between the structure of the Human Estrogen Receptor alpha Ligand Binding Domain (3ERT:A), CAA05410.2 (LBDG5) and AAK49953.1 (LBDG13). Genome Threader™ and Inpharmatica PSI-BLAST have identified residues 394-604 of CAA05410.2 (LBDG5) as adopting the structure of residues 15-247 of the Human Estrogen Receptor alpha Ligand Binding Domain (3ERT:A), double-headed arrow at the top of Figure 18. The relationships between (3ERT:A), CAA05410.2 (LBDG5) and AAK49953.1 (LBDG13) are also presented in the form of a multiple alignment (Figure 19). This alignment has been generated by first taking the Genome Threader alignment of 3ERT:A with CAA05410.2 (LBDG5), and then adding the sequence of AAK49953.1 (LBDG13) to the alignment on the basis of it's sequence identity to

CAA05410.2 (LBDG5). By reference to the alignment it is possible to map the domain boundaries of the predicted Ligand Binding Domain of CAA05410.2 (LBDG5) onto AAK49953.1 (LBDG13), these domain boundary residues are marked by grey boxes in Figure 19. This identifies that residues 117-317 of AAK49953.1 (LBDG13) correspond to the residues 394-604 of CAA05410.2 (LBDG5) that are predicted to adopt the structure of a Ligand Binding Domain by Genome Threader and Inpharmatica PSI-BLAST (see also Figure 18). Thus we annotate residues 117-317 of AAK49953.1 (LBDG13) as adopting the structure of a Nuclear Hormone Receptor Ligand Binding Domain on the basis of sharing high sequence homology with residues 394-604 of CAA05410.2 (LBDG5).

5 10 In order to view the public domain annotation of AAK49953.1 (LBDG13) the InterPro secondary database is queried with AAK49953.1 (LBDG13; Figure 30). It can be seen from Figure 30 that no matches are found to InterPro. Returning no matches demonstrates that AAK49953.1 (LBDG13) is unidentifiable as containing a Nuclear Hormone Receptor Ligand Binding Domain using InterPro.

15 20 In order to view what is known in the public domain secondary databases, the NCBI Conserved Domain Database (CDD) is queried with AAK49953.1 (LBDG13; Figure 31). CDD returns only 1hit to the PFAM profile PF02562, PhoH. This profile match may be a false positive since only 32.3% of the PhoH profile matches AAK49953.1 (LBDG13). Irrespective of whether this PhoH match is a true or false positive, PhoH has no connection to the Nuclear Hormone receptor Ligand Binding Domain family and thus CDD does not annotate AAK49953.1 (LBDG13) as containing a Nuclear Hormone Receptor Ligand Binding Domain using CDD.

25 30 NCBI provides a public domain PSI-Blast server. Querying NCBI PSI-Blast with AAK49953.1 (LBDG13) through 10 positive iterations fails to annotate any region of AAK49953.1 (LBDG13) as having a relationship to any known Nuclear Hormone Receptor Ligand Binding Domains (note that NCBI PSI-Blast cannot provide data on negative iterations because no all-by-all calculation is performed). Figure 32 shows the graphical display of NCBI PSI-Blast results for AAK49953.1 (LBDG13). Figure 33 lists these NCBI PSI-BLAST hits (marked by lines). None of these sequences have been annotated in the public domain as containing a Nuclear Hormone Receptor Ligand

Binding Domain. Thus NCBI PSI-Blast does not annotate AAK49953.1 (LBDG13) as having a relationship to any known Nuclear Hormone receptor Ligand Binding Domains.

The National Centre for Biotechnology Information (NCBI) GenBank protein database is viewed to examine if there is any further information that is known in the public domain 5 relating to AAK49953.1 (LBDG13). This is the U.S. public domain database for protein and gene sequence deposition (Figure 33). AAK49953.1 was cloned by a group of scientists at the Indiana University School of Medicine, USA. The authors annotate AAK49953.1 (LBDG13) as a protein that is similar to CAC17441. CAC17441 is annotated as being similar to p65 protein, but as discussed earlier, p65 (CAA05410.2 (LBDG5)) is not 10 annotated in the public domain as containing a Nuclear Hormone Receptor Ligand Binding Domain. Thus the (NCBI) GenBank does not in any way annotate AAK49953.1 (LBDG13) as containing a Nuclear Hormone Receptor Ligand Binding Domain or suggest that the protein has this function.

There is no further public domain annotation for AAK49953.1 (LBDG13). The public 15 domain information for this protein does not annotate it as containing a Nuclear Hormone Receptor Ligand Binding Domain. Therefore using all public domain annotation tools, AAK49953.1 (LBDG13) is not annotated as containing a Nuclear Hormone Receptor Ligand Binding Domain.

To summarise we annotate residues 117-317 of AAK49953.1 (LBDG13) as adopting the 20 structure of a Nuclear Hormone Receptor Ligand Binding Domain on the basis of sharing high sequence homology with residues 394-604 of CAA05410.2 (LBDG5) which Genome Threader and Inpharmatica PSI-BLAST predict will adopt the structure of a Nuclear Hormone Receptor Ligand Binding Domain.

Sequence Listing

SEQ ID NO:1 (AJ002425.2 Nucleotide sequence for CAA05410.2 (LBDG5) protein, coding region start position =61, coding region finish position=1905)

5 1 ggggcttatt aacgggtggta tctggtagag ccccccataaaa gtaccaccgt ggggaagaca
61 atgggaccgg aaaacgagag ccccggtgggg ggggagatgt gggggagagt gtgtgtttc
121 agcgctagac aaaagagacc gcgatgcagc actttaaata gtcttctcgt tgctagaaca
181 tgctataaca tgctggcgct ggtggattgg atcgagcagg atttcaagag cacttggcga
241 aaggctcacg tttccctccg cgggttcctc gatttcgagc tggggagacg aaaagaggtt
10 301 gcccgggctt ttctgggtgg ggataccaga cctgacccca aaaagcccgcg ggggggggtcc
361 aaaaagaacg tggaggtgta tgatgatgac gtggctctc aggctgcgga cagccccggg
421 aaacgcatgg ccccgaaagg gacatttaga gataaggaca aatttgaagg gctgtttaag
481 ctcggggcgc tggggcaaa aaaagccttg accccagcat tttcttggc ccccaacagg
541 gggtcgcctc tacacgcccc ttatgggat gaaatcctct acaaggttga atccggggcc
15 601 gtcaacattt gcgagggggg caaaagaggc gtgggatcc acccccccaga taactacggc
661 gatacccttg atgaaaatct tggccttccc caaaaaattt tgattaaggt gaagcccaa
721 accgaggaag ccaacacttg gttaaggcag gatctgaaaa atcataacag tgcaaaggag
781 gccggggct ccgacgagat taaaaccttt gtgacaggat gtaaaaaaga tgggcataagt
841 gggcgtaaaa atatgaccac acatgacaga aattcaaaaa agtggcaacg gttaaaacttg
20 901 tcccttatgg ctccttccca gctagattct aggggcggac gcgcggggcc cggcgcggga
961 gcgcggcggcc tggccttggt gtgtgaggac tatgccagct gttcaaacac ctgtgtctgg
1021 tcctgtgaag cctacaaggt cttcttcgc cgaagtcaaa gttcacaga tccagcctgt
1081 ttcacaaaacg attgcaacat ctctaagaat aggtctaagt cttgcccagc ttgcctcctc
1141 cggtgcctgc accctagcat taatgagatc cgaaaagaca agcgagcagc gctgaaagtg
25 1201 cgagacaacg ttgggtgaaga ggtggatatg accggcccta gctggacactg cctgaagcta
1261 ctttttcag atggggaaaa agtgataaccc agatgggccc atgaactccc agggatcaag
1321 gggggccggc aggcaaaaaca gcagtcacac cgaggaagcc ccatccccaa aaacaggaaa
1381 ggttggcccc ccggacatgt cctgtcaaatt gacggcggag ctggggcag ggtatggaaa
1441 aaaaaatcct gtaaaaccaat tcgcccagaa ggccccaaat ggtggatcg gctgaatgaa
30 1501 tctacacctt tgaaaaatgggt gtctcgagcc aacaagagt tagggaaagg aggccaccagg
1561 gggaggatt tcatcaagca cccacaccc tttaagtttgc cagcagatcc tcaggacaag
1621 cactggctgg ctgagcagca tcatatgcgg gcaacaggag gaaagatggc gtaccccttc
1681 attgaggaag acatcggca gcatcatggc caggggttcc cagttatgct tctcaagatt
1741 agccatatta ggcacatggt tggggagatg gctcattgct tgtacgacat gaaagaaaaag
35 1801 aagtttggc tgccatcctg gaagggttgc aagttgggaa aatacgttgc gacactacgg
1861 acagaaaaag agcatcgatc tgctgaagca agtccccaga cctgactttc ccggccccggc
1921 tgaggccatc atggggatgc ggtcttagttg gctcttagca gcatcaagct gtacatgagc
1981 tagttttagt tgactcactg cagagcccc cagactggct tggggcttc tttctaaatgt

2041 tatttggata agaaggcaatt aaacaaggttt gtaatttaaa aaaaaaaaaa aaaaa

SEQ ID NO:2 (Protein CAA05410.2; LBDG5)

1 mgpenespv gemwgrvcvf sargkrprcs tlnsllvart cynmlalvdw ieqdfkstwr
 5 61 kahvslrgfl dfelgrrkev agafllggdtr pdpkprrggs kknvevyddd vgsqaadspg
 121 krmapkgtfr dkdkfeglfk lgalvakkal tpafscspnr gsplhahygd eilykvesgp
 181 vniceggkrg vgihppdnyg dtldenlglp qkivikvkpq teeantwlrq dlknhnsake
 241 aggsdeiktf vtgcckdghs grknmtthdr nskkwqrvn1 slmaslqls rggragprrg
 301 arrlclvced yascnntcvw sceaykvffr rsqsftdpac ftndcniskn rskscpacll
 10 361 rclhpsinei rdkdkraalkv rdnvgeevdm tggpswtclkl lfsdgekvip rlghelpgik
 421 ggrqakqqsh rgspipknrk gwppghvlsn dggaggrvwk kksckpirre gpkwwdrlne
 481 stplfwgsra nkslgkggtr grifikhphl fkfaadpqdk hwlaeqhhmr atggkmayll
 541 ieeditgqhhg qgfpvml1ki shirhmvggv ahclydmkek kfvlpswkve klgkyvetlr
 601 tekehraaea spqt

15

SEQ ID NO:3 (Nucleotide coding sequence for LBDG10, CAA05409.2)

1 atggaccccg agaacgtgag agccgagggc gggacatga gggagaaagt gagactgtcc
 61 agcgctagac agagactgcg atgcagcact aagaatagtc ttctcggtgc tagaacatgc
 121 gtgctgaggg tggtcgattt gatcgagcag gatctcaaga gcacttgcga aaaggctcac
 181 gttccctcc gcgggtaccc cgatttccag ctgcaatgcg ggagacgaaa agaggttgcc
 241 gggctctgg gtggggatac cagacctgac cccaaagaagc cgcggggggg ctccaaaaag
 301 aacgtggagg tgtatggta tgacgttaggc tctcaggctg cggacagccc gaggaaacag
 361 ctggccgcga aaggacatt cagagataag gacaaaattt aagcgctgtt caagctcggg
 421 gagctgggtgg ctaaaaaaagc cttgtcctca gcaattactt gttcccaa cagcgtgtcg
 481 cctctacacg cccattatgg gatgaaatc ctctacaagg atgaatccgg gcttgtcaac
 541 attagcgagg gtggcaaaag aggcgtggag atccacccccc cagataactt cggcatcaca
 601 accctttagt aagatcttgg ctttcccaa ataattgtga ttaacgtgaa gcctcaaacc
 661 gaggaaagcca acacttggta tagacaggat ctgaaatatc ataacagtgc aaacgaggcc
 721 gggtaactccg acgagaataa aacctttgtg agaggatgta gaaaagatgg gcatagtgag
 781 cgtaacaata tgaccacagg tgacagaaat tcaaaaaagg cccagccgt aaacttctcc
 841 ctcatggcat ccctcgccct ggattctagg ggcaaagccg cggggccccc ggcggagcgc
 901 aggcgcctgt gcctgggtgt tgaggactat gccagctgtt caaacacctg tgtctggtcc
 961 tgtgaagcct acaaggctt ctttcgcccga agtcaaagtt tcacagatcc agcctgtttc
 1021 acaaacgatt gcaacatctc taagaataga tctaagtctt gcccagccgt cctcctccgt
 1081 tgcctgcagc ctagcatcaa tgagatccga aaagacaagc gagcagcgct gaatgtgcga
 1141 gacaacgttg gtgaagaggt ggatatgacc ggtccttagct ggacctgcct gaagctactc
 1201 ttttcagatg gagaaaaaagt gataccaga ttggccatg aacttccagg gatcaagcgt

1261 ggccggcagg cagaagagga gtcaccgcga ggaagccccca ttcccaaaaaa gaggaaaggt
 1321 tggcctcctg gacatgtcct gtcaaattgac cgccgcagctg ctggcacgggt atggaaacca
 1381 aaatcctgtg aaccaattcg ccgagaaggc cccaaatgggg acgctcggct gaatgaatct
 1441 accaccccttgc ttttggggtc tcgagccaaac aaggccttag ggaaggggagg caccagaggg
 5 1501 aggatttaca tcaagcaccc acaccccttt aagtatgcag cagatcctca ggacaagcac
 1561 tggctggctg agcagcttca tatgcgggca acaggaggaa agatggcgta ccttctcatt
 1621 gaggaagaca tcgagcagca tcatggccag cggttcgcag ttttccttct caagattagc
 1681 catattaggc acatgggtga gggagtggtc cattgcttgc acgacatgaa agttatccag
 1741 tttgttctgc catcctggaa ggttgagaag ttgcggaaat acgtggagac actacggaca
 10 1801 gaaaatgagc atcgtgctgc tgaagcaagt ccccagaccc tga

SEQ ID NO:4 (Protein coding sequence for LBDG10, CAA05409.2)

1 mdpenvraeg gdmrekvrls sargrlrcst knsllvartc virlvdwieq dlkstcekah
 15 61 vslrgyldfq lqcgrrkeva galggdtrpd pkkprggsskk nvevygddvg sqaadspkq
 121 laakgtfrdk dkiaealfklg elvakkalss aitwfpsvs plhahygdei lykdesglvn
 181 iseggkrgve ihppdnfgit tldedlgfpq iivinvkpqt eeantwyrqd lkyhnsanea
 241 gysdenktfv rgcrkdgħse rnmmttgħdrn skkaqpvnfs imaslaldsr gkaagprrga
 301 rrlclvcedy ascsntcvws ceaykvffrr sqsftdpacf tndcnisknr skscpacllr
 20 361 clqpsineir kdkraalnvr dnvgeevdm t gpswtclkll fsdgekvipr lahelpgikr
 421 grqaeeshr gspipkkkrkg wppghvlsnd raaagtvwkp kscepirreg pkwdarlnes
 481 ttfvlgsran kalgkggtrg riyikhphlf kyaadpqdkh wlaeqlhmra tggkmaylli
 541 eedieqhhgq rfavflkkis hirhmvegva hclydmkviq fvlpswkvek lrkyvetlrt
 601 enehraaaeas pqt

25

SEQ ID NO:5 (Nucleotide coding sequence for BAB62888.1 (LBDG11))

1 atgggagcca ctggcgacgc cgagcagccg cggggaccca gcggggccga gaggggcggc
 61 ttggagctgg gggatgcggg cgccagcgggg cagctggttc ttacgaaccc ttggaaacata
 121 atgataaaagc accggcaggt gcagcggagg ggccgcgcgt cacagatgac aacaagttc
 30 181 acagatcctg ccatctccat ggatctcctc cgagctgtcc tgcagccag catcaacgag
 241 gagatccaga ctgtcttcaa caagtacatg aagttcttcc agaaggcagc actgaacgtg
 301 cgagacaatg ttggggagga ggtggacgca gagcagctga tccaggaagc ctgtcggagc
 361 tgcctggagc aggctaaact gctctttca gatggagaaa aagtaatacc cagattgacc
 421 catgagcttc caggaataaa gcgtggccgt cagggcagaag aagaatgtgc ccatcgagga
 35 481 agcccccttc ctaaaaagag gaaaggacgg cctcctggac acatcctgtc aagcgaccgg
 541 gcagccgccc gcatggatg gaaaccaaaa tcctgtgaac caattcgccg ggaaggcccc
 601 aagtgggacc cagctcgctt gaatgaatct accaccccttgc tggatggatc tcgagccaa

661 aaagccctgg ggatgggggg caccagagga agaatttaca tcaaggaccc acaccccttt
 721 aagtatgcag ctgaccccca ggataagcac tggctggctg agcagcatca catgcgggca
 781 acagggggca agatggccta cctcctcatc gaggaggaca tccgggaccc tgcggccagt
 841 gatgattaca gaggatgcct ggatctgaag ctagaggaat tgaatcctt tgccttaccc
 5 901 tcctggatgg tggagaagat gagaaagtat atggagacac tacggacaga gaatgagcat
 961 cgtgctgtt aagcacctcc acagacctga

SEQ ID NO:6 (Protein coding sequence for BAB62888.1 (LBDG11))

1 mgatgdaeqp rgpsgaergg lelgdagaag qlvltnpwni mikhrqvqrr grrsqmstsf
 10 61 tdpaismdll ravlqpsine eiqtvfnkym kffqkaalnv rdnvgeevda eqliqeacrs
 121 cleqakllfs dgekviprlt helpgikrgr qaeecahrg splpkkrkgr ppghilssdr
 181 aaagmvwkpk scepirregp kwdparlnes ttfvlgsran kalgmggtrg riyikhphlf
 241 kyaadpqdkh wlaeqhhmra tggkmaylli eedirdlaas ddyrgcldlk leelksfvlp
 301 swmvekmrky metlrteneh raveappqt

15

SEQ ID NO:7 (Nucleotide coding sequence for AAH03486.1 (LBDG12))

1 atgggggcga ctggcgacac cgagcagccg cggggcccccgc gcggggcggga gcgaggtggc
 61 ctggagctgg gcgacgcggg cgccggcgggc cagccggttc tcacgaaccc ttggaacata
 121 atgataaaac atcggcaggt gcagcgaagg ggccgcgcgat ctcagatgac cacaagttc
 20 181 acagacccag ccattcttat ggatctcctc cgtgctgtcc tgcagcctag catcaatgag
 241 gagatccagg gtgtcttcaa caagtacatg aagttcttcc agaaggcagc gctgaatgtg
 301 cgagacaatg ttgggaaaga agtggacgca gagcagttga ttcaggaggc ctgccgcagc
 361 tgcctggagc aggcaaagct gctctttct gatggagaga aagtataacc cagattggcc
 421 catgagcttc cagggatcaa gcggggccgg caagcagaag aggagtccca ccgaggaagc
 25 481 cccattccca aaaagaggaa aggtcggcct cctggacacg tcctgtcaaa tgaccgcga
 541 gctgctggca tggtatggaa acaaaaatcc tgtgaaccaa ttcgcccaga aggccccaaag
 601 tgggacccag ctcggctgaa tgaatctacc acctttgttt tggggtctcg agccaaacaag
 661 gccttgggaa tgggaggcac cagagggaga atctacatca agcaccacaca ccttttaag
 721 tatgcagcag atcctcagga caagcactgg ctggctgagc agcatcacat gcgccaaaca
 30 781 ggccggaaaga tggcgtaccc tctcattgaa gaagacatcc gggacttggc tgccagcgat
 841 gactacagag gatgcttggc cctgaagttg gaggagctga aatcctttgt tctgccatcc
 901 tggatggttt agaagatgcg gaaatacatg gagacactgc ggacagaaaa tgagcaccgc
 961 gctgcggaag cgcctccca gacctga

35 SEQ ID NO:8 (Protein coding sequence for AAH03486.1 (LBDG12))

1 mgatgdteqp rgpggaergg lelgdagaag qpvltnpwni mikhrqvqrr grrsqmstsf
 61 tdpaismdll ravlqpsine eiqgvfnkym kffqkaalnv rdnvgeevda eqliqeacrs

121 cleqakllfs dgekviprla helpgikrgr qaeeshrgs pipkrkgrp pghvlsndra
 181 aagmvwkqks cepirregpk wdpalnest tfvlgsrank alggggtrgr iyikhphlfk
 241 yaadpqdkhw laeqhhmrat ggkmayllie edirdlaasd dyrgcldlkl eelksfvlp
 301 wvekmrkym etlrtnehr aaeappqt

5

SEQ ID NO:9 (Nucleotide coding sequence for AAK49953.1 (LBDG13))

1 atggggcga ctggcgacac cgggggnccc cggccgggga cggatcaag gcggcctgga
 61 aacgtggca acgcgggcgc ggcgggcccag ccggttctca cgaacccttg gaacataatg
 121 atcaaaccctc ggcaggtgca gcaagagggc cgccgatctc agatgaccac aagtttcaca
 10 181 gatccagcca tctctatgga tctcctccgt gctgtcctgc agcctagcat caatgaggag
 241 atccagagtg tcttcaacaa gtacatgaag ttcttccaga aggcagcgct gaatgtgcga
 301 gacaacgttg gtgaagaggt ggatgcagaa cagttgattc aggaggcctg ccgcagctgc
 361 gtggagcagg caaagctact ctttcagat ggagaaaaag tgataccag attggccat
 421 gaacctccag ggatcaagcg tggccggcag gcagaagagg agtcccaccc agaagcccc
 15 481 ttcccaaaaa gaggaaaggt cggccttcct ggacatgtcc tgtcaaatga ccgcgcagct
 541 gctggcatgg tatggaaacc aaaatcctgt gaaccaattc gccgagaagg ccccaagtgg
 601 gaccagctc ggctgaatga atctaccacc tttgtttgg ggtctcgagc caacaaggcc
 661 ttagggatgg gaggcaccag agggaggatt tacatcaagc acccacacct ctttaagtat
 721 gcagcagatc ctcaggacaa gcactggctg gctgagcagc atcatatgct ggcaacagga
 20 781 ggaaagatgg cgtacccctt cattgaggaa gacatccgag acttggctgc cagcgtatgc
 841 tacagaggat gcttggaccc gaagttggag gagctgaagt ctttggatggatggatgg
 901 atgggtgaga agatgcggaa atacatggag acactacgga cagaaaatga gcatcgtgct
 961 gctgaagcaa ctccccagac ctga

25 SEQ ID NO:10 (Protein coding sequence for AAK49953.1 (LBDG13))

1 mgatgdtggp rpgtesrrpg nvgnagaagg pvltnpwnim ikprqvqrrg rrsqmttsft
 61 dpaismdllr avlqpsinee iqsvfnkymk ffqkaalnvr dnvgeevdae qliqeacrsc
 121 veqakllfsd gekviprlah elpgikrgrq aeeeshreap fpkrgkvglp ghvlsndraa
 181 agmvwkpksc epirregpkw dparlnestt fvlgsranka lggggtrgrri yikhphlfky
 30 241 aadpqdkhw aeqhhmratg gkmaylliee dirdlaasd yrgcldlkle elksfvlp
 301 wvekmrkyme tlrtnehr aeatpqt

CLAIMS

1. A polypeptide, which polypeptide:

5 (i) comprises or consists of the amino acid sequence as recited in SEQ ID NO:2;

(ii) is a fragment thereof having activity as a Nuclear Hormone Receptor Ligand Binding Domain or having an antigenic determinant in common with the polypeptide of (i); or

(iii) is a functional equivalent of (i) or (ii).

10 2. A polypeptide which is a fragment according to claim 1(ii), which includes the Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG5 polypeptide, said Nuclear Hormone Receptor Ligand Binding Domain region being defined as including between residues 394 and 604 inclusive of the amino acid sequence recited in SEQ ID NO:2 and possesses activity as a Nuclear Hormone Receptor Ligand Binding Domain.

15 3. A polypeptide which is a functional equivalent according to claim 1(iii), is homologous to the amino acid sequence as recited in SEQ ID NO:2, and has activity as a Nuclear Hormone Receptor Ligand Binding Domain.

20 4. A polypeptide according to claim 3, wherein said functional equivalent is homologous to the Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG5 polypeptide.

25 5. A fragment or functional equivalent according to any one of claims 1-4, which has greater than 80% sequence identity, preferably greater than 85%, 90%, 95%, 98% or 99% sequence identity, with an amino acid sequence as recited in SEQ ID NO:2, or with a fragment thereof that possesses Nuclear Hormone Receptor Ligand Binding Domain activity, said sequence identity being determined using BLAST version 2.1.3 using the default parameters specified by the NCBI (the National Center for Biotechnology Information; <http://www.ncbi.nlm.nih.gov/>) [Blosum 62 matrix; gap open penalty=11 and gap extension penalty=1].

6. A functional equivalent according to any one of claims 1-5, which exhibits significant structural homology with a polypeptide having the amino acid sequence given in SEQ ID NO:2, or with a fragment thereof that possesses activity as a Nuclear Hormone Receptor Ligand Binding Domain.

5 7. A functional equivalent according to any one of claims 2-6, which

- i) comprises or consists of the amino acid sequence as recited in SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8 or SEQ ID NO:10;
- 10 ii) is a fragment thereof having activity as a Nuclear Hormone Receptor Ligand Binding Domain or having an antigenic determinant in common with the polypeptide of (i).

8. A functional equivalent according to claim 7(ii), which includes

- i) the Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG10 polypeptide, said Nuclear Hormone Receptor Ligand Binding Domain region being defined as including between residues 413 and 603 inclusive of the amino acid sequence recited in SEQ ID NO:4 and possesses activity as a Nuclear Hormone Receptor Ligand Binding Domain; or
- 15 ii) the Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG11 polypeptide, said Nuclear Hormone Receptor Ligand Binding Domain region being defined as including between residues 118 and 319 inclusive of the amino acid sequence recited in SEQ ID NO:6 and possesses activity as a Nuclear Hormone Receptor Ligand Binding Domain; or
- 20 iii) the Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG12 polypeptide, said Nuclear Hormone Receptor Ligand Binding Domain region being defined as including between residues 118 and 318 inclusive of the amino acid sequence recited in SEQ ID NO:8 and possesses activity as a Nuclear Hormone Receptor Ligand Binding Domain; or
- 25 iv) the Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG13 polypeptide, said Nuclear Hormone Receptor Ligand Binding Domain region being defined as including between residues 117 and 317

inclusive of the amino acid sequence recited in SEQ ID NO:10 and possesses activity as a Nuclear Hormone Receptor Ligand Binding Domain.

9. A fragment as recited in claim 1, 2, 5, or 7 having an antigenic determinant in common with the polypeptide of claim 1(i) or claim 7(i) which consists of 7 or more (for example, 8, 10, 12, 14, 16, 18, 20 or more) amino acid residues from the sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO: 8 or SEQ ID NO:10 respectively.
10. A polypeptide or polypeptide fragment according to any one of the preceding claims that is in the form of a dimer complex.
- 10 11. A dimer complex according to claim 10, which is a homodimer.
12. A dimer complex according to claim 11, which is a heterodimer.
13. A dimer complex according to claim 12, wherein said polypeptide or fragment is complexed with a polypeptide comprising a Nuclear Hormone Receptor Ligand Binding Domain.
- 15 14. A dimer complex according to claim 13, wherein the LBDG5 polypeptide or a LBDG10 polypeptide or fragment thereof is complexed with the estrogen receptor α ligand binding domain, or the estrogen receptor β ligand binding domain.
- 15 15. A purified nucleic acid molecule which encodes a polypeptide or dimer complex according to any one of the preceding claims.
- 20 16. A purified nucleic acid molecule according to claim 15, which has the nucleic acid sequence as recited in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7 or SEQ ID NO:9 or is a redundant equivalent or fragment thereof.
17. A fragment of a purified nucleic acid molecule according to claim 15 or claim 16, which comprises between nucleotides 1240 and 1872 of SEQ ID NO:1, between nucleotides 1237 and 1809 of SEQ ID NO:3, between nucleotides 352 and 957 of SEQ ID NO:5, between nucleotides 352 and 954 of SEQ ID NO:7, between nucleotides 349 and 951 of SEQ ID NO:9, or is a redundant equivalent thereof.

18. A purified nucleic acid molecule which hybridizes under high stringency conditions with a nucleic acid molecule according to any one of claims 15-17.
19. A vector comprising a nucleic acid molecule as recited in any one of claims 15-18.
20. A host cell transformed with a vector according to claim 19.
- 5 21. A ligand which binds specifically to, and which preferably inhibits the Nuclear Hormone Receptor Ligand Binding Domain activity of, a polypeptide according to any one of claims 1-10 or a dimer according to any one of claims 11-14.
22. A ligand according to claim 21, which is an antibody.
- 10 23. A compound that either increases or decreases the level of expression or activity of a polypeptide according to any one of claims 1-10.
24. A compound according to claim 23 that binds to a polypeptide according to any one of claims 1-10 without inducing any of the biological effects of the polypeptide.
25. A compound according to claim 23 or claim 24, which is a natural or modified substrate, ligand, enzyme, receptor or structural or functional mimetic.
- 15 26. A polypeptide according to any one of claims 1-8, a nucleic acid molecule according to any one of claims 15-18, a vector according to claim 19, a host cell according to claim 20, a ligand according to claim 21 or claim 22, or a compound according to any one of claims 23-25, for use in therapy or diagnosis of disease.
27. A method of diagnosing a disease in a patient, comprising assessing the level of expression of a natural gene encoding a polypeptide according to any one of claims 1-10, or assessing the activity of a polypeptide according to any one of claims 1-10, in tissue from said patient and comparing said level of expression or activity to a control level, wherein a level that is different to said control level is indicative of disease.
28. A method according to claim 27 that is carried out *in vitro*.
- 25 29. A method according to claim 27 or claim 28, which comprises the steps of: (a) contacting a ligand according to claim 14 or claim 15 with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and (b) detecting said complex.

30. A method according to claim 27 or claim 28, comprising the steps of:

a) contacting a sample of tissue from the patient with a nucleic acid probe under stringent conditions that allow the formation of a hybrid complex between a nucleic acid molecule according to any one of claims 15-18 and the probe;

5 b) contacting a control sample with said probe under the same conditions used in step a); and

c) detecting the presence of hybrid complexes in said samples;

wherein detection of levels of the hybrid complex in the patient sample that differ from levels of the hybrid complex in the control sample is indicative of disease.

10 31. A method according to claim 27 or claim 28, comprising:

a) contacting a sample of nucleic acid from tissue of the patient with a nucleic acid primer under stringent conditions that allow the formation of a hybrid complex between a nucleic acid molecule according to any one of claims 15-18 and the primer;

15 b) contacting a control sample with said primer under the same conditions used in step a); and

c) amplifying the sampled nucleic acid; and

d) detecting the level of amplified nucleic acid from both patient and control samples;

20 wherein detection of levels of the amplified nucleic acid in the patient sample that differ significantly from levels of the amplified nucleic acid in the control sample is indicative of disease.

32. A method according to claim 27 or claim 28 comprising:

a) obtaining a tissue sample from a patient being tested for disease;

25 b) isolating a nucleic acid molecule according to any one of claims 15-18 from said tissue sample; and

c) diagnosing the patient for disease by detecting the presence of a mutation which is associated with disease in the nucleic acid molecule as an indication of the disease.

33. The method of 31, further comprising amplifying the nucleic acid molecule to form an amplified product and detecting the presence or absence of a mutation in the 5 amplified product.

34. The method of either claim 32 or 33, wherein the presence or absence of the mutation in the patient is detected by contacting said nucleic acid molecule with a nucleic acid probe that hybridises to said nucleic acid molecule under stringent conditions to form a hybrid double-stranded molecule, the hybrid double-stranded molecule having an unhybridised portion of the nucleic acid probe strand at any portion corresponding to 10 a mutation associated with disease; and

detecting the presence or absence of an unhybridised portion of the probe strand as an indication of the presence or absence of a disease-associated mutation.

35. A method according to any one of claims 27-34, wherein said disease is a disease in 15 which Nuclear Hormone Receptor Ligand Binding Domains are implicated, such as cell proliferative disorders, including neoplasm, melanoma, lung, colorectal, breast, uterus, prostate, pancreas, head and neck and other solid tumours, myeloproliferative disorders, such as leukemia, non-Hodgkin lymphoma, leukopenia, thrombocytopenia, angiogenesis disorder, Kaposi's sarcoma, autoimmune/inflammatory disorders, 20 including allergy, inflammatory bowel disease, arthritis, psoriasis and respiratory tract inflammation, asthma, and organ transplant rejection, cardiovascular disorders, including hypertension, hypotension, oedema, angina, atherosclerosis, thrombosis, sepsis, shock, reperfusion injury, heart arrhythmia, and ischemia, neurological disorders including, central nervous system disease, Alzheimer's disease, Parkinson's 25 disease, brain injury, stroke, amyotrophic lateral sclerosis, anxiety, depression, and pain, cognition enhancement, learning and memory enhancement, developmental disorders, metabolic disorders including diabetes mellitus, osteoporosis, lipid metabolism disorder, hyperthyroidism, hyperparathyroidism, thyroid hormone resistance syndrome, hypercalcemia, hypocalcaemia, hypercholesterolemia, 30 hyperlipidemia, and obesity, renal disorders, including glomerulonephritis,

renovascular hypertension, blood disorders including hemophilia, dermatological disorders, including, cellulite, acne, eczema, psoriasis and wound healing, scarring, negative effects of aging, fertility enhancement, contraception, pregnancy termination, progesterone antagonism, hormone replacement therapies, steroid hormone-like mediated hair characteristics, immunomodulation, AIDS, vision disorders, glucocorticoid resistance, mineralocorticoid resistance, androgen resistance, pseudohypoaldosteronism, spinal/bulbar muscular atrophy, extraskeletal myxoid chondrosarcomas, adrenal insufficiency, sexual reversal, infections including viral infection, bacterial infection, fungal infection and parasitic infection, cancer, 5 particular cancers originating from estrogen-responsive tissues, including breast, uterus and prostate, myeloproliferative disorders, such as leukemia, hypertension, hypotension, fertility enhancement, contraception, pregnancy termination, progesterone antagonism, wound healing, scarring, obesity, dermatological disorders including cellulite, estrogen-mediated hair characteristics, central nervous system 10 disorders, Alzheimer's disease, cognition enhancement, learning and memory enhancement, immunomodulation and osteoporosis.

15

36. Use of a polypeptide according to any one of claims 1-10 as a Nuclear Hormone Receptor Ligand Binding Domain.
37. Use of a nucleic acid molecule according to any one of claims 15-18 to express a 20 protein that possesses Nuclear Hormone Receptor Ligand Binding Domain activity.
38. A pharmaceutical composition comprising a polypeptide according to any one of claims 1-10, a nucleic acid molecule according to any one of claims 15-18, a vector according to claim 19, a host cell according to claim 20, a ligand according to claim 21 or claim 22, or a compound according to any one of claims 23-25.
- 25 39. A vaccine composition comprising a polypeptide according to any one of claims 1-10 or a nucleic acid molecule according to any one of claims 15-18.
40. A polypeptide according to any one of claims 1-10, a nucleic acid molecule according to any one of claims 15-18, a vector according to claim 19, a host cell according to claim 20, a ligand according to claim 21 or claim 22, a compound according to any 30 one of claims 23-25, or a pharmaceutical composition according to claim 32 for use

in the manufacture of a medicament for the treatment of a disease in which Nuclear Hormone Receptor Ligand Binding Domains are implicated, such as cell proliferative disorders, including neoplasm, melanoma, lung, colorectal, breast, uterus, prostate, pancreas, head and neck and other solid tumours, myeloproliferative disorders, such as leukemia, non-Hodgkin lymphoma, leukopenia, thrombocytopenia, angiogenesis disorder, Kaposi's sarcoma, autoimmune/inflammatory disorders, including allergy, inflammatory bowel disease, arthritis, psoriasis and respiratory tract inflammation, asthma, and organ transplant rejection, cardiovascular disorders, including hypertension, hypotension, oedema, angina, atherosclerosis, thrombosis, sepsis, shock, reperfusion injury, heart arrhythmia, and ischemia, neurological disorders including, central nervous system disease, Alzheimer's disease, Parkinson's disease, brain injury, stroke, amyotrophic lateral sclerosis, anxiety, depression, and pain, cognition enhancement, learning and memory enhancement, developmental disorders, metabolic disorders including diabetes mellitus, osteoporosis, lipid metabolism disorder, hyperthyroidism, hyperparathyroidism, thyroid hormone resistance syndrome, hypercalcemia, hypocalcaemia, hypercholesterolemia, hyperlipidemia, and obesity, renal disorders, including glomerulonephritis, renovascular hypertension, blood disorders including hemophilia, dermatological disorders, including, cellulite, acne, eczema, psoriasis and wound healing, scarring, negative effects of aging, fertility enhancement, contraception, pregnancy termination, progesterone antagonism, hormone replacement therapies, steroid hormone-like mediated hair characteristics, immunomodulation, AIDS, vision disorders, glucocorticoid resistance, mineralocorticoid resistance, androgen resistance, pseudohypoaldosteronism, spinal/bulbar muscular atrophy, extraskeletal myxoid chondrosarcomas, adrenal insufficiency, sexual reversal, infections including viral infection, bacterial infection, fungal infection and parasitic infection, cancer, particular cancers originating from estrogen-responsive tissues, including breast, uterus and prostate, myeloproliferative disorders, such as leukemia, hypertension, hypotension, fertility enhancement, contraception, pregnancy termination, progesterone antagonism, wound healing, scarring, obesity, dermatological disorders including cellulite, estrogen-mediated hair characteristics, central nervous system

disorders, Alzheimer's disease, cognition enhancement, learning and memory enhancement, immunomodulation and osteoporosis.

41. A method of treating a disease in a patient, comprising administering to the patient a polypeptide according to any one of claims 1-10, a nucleic acid molecule according to any one of claims 15-18, a vector according to claim 19, a host cell according to claim 20, a ligand according to claim 21 or claim 22, a compound according to any one of claims 23-25, or a pharmaceutical composition according to claim 32.
5
42. A method according to claim 41, wherein, for diseases in which the expression of the natural gene or the activity of the polypeptide is lower in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, vector, ligand, compound or composition administered to the patient is an agonist.
10
43. A method according to claim 42, wherein, for diseases in which the expression of the natural gene or activity of the polypeptide is higher in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, vector, ligand, compound or composition administered to the patient is an antagonist.
15
44. A method of monitoring the therapeutic treatment of disease in a patient, comprising monitoring over a period of time the level of expression or activity of a polypeptide according to any one of claims 1-10, or the level of expression of a nucleic acid molecule according to any one of claims 15-18 in tissue from said patient, wherein altering said level of expression or activity over the period of time towards a control level is indicative of regression of said disease.
20
45. A method for the identification of a compound that is effective in the treatment and/or diagnosis of disease, comprising contacting a polypeptide according to any one of claims 1-10, a nucleic acid molecule according to any one of claims 15-18, or a host cell according to claim 20 with one or more compounds suspected of possessing binding affinity for said polypeptide or nucleic acid molecule, and selecting a compound that binds specifically to said nucleic acid molecule or polypeptide.
25

46. A kit useful for diagnosing disease comprising a first container containing a nucleic acid probe that hybridises under stringent conditions with a nucleic acid molecule according to any one of claims 15-18; a second container containing primers useful for amplifying said nucleic acid molecule; and instructions for using the probe and primers for facilitating the diagnosis of disease.

5 47. The kit of claim 46, further comprising a third container holding an agent for digesting unhybridised RNA.

48. A kit comprising an array of nucleic acid molecules, at least one of which is a nucleic acid molecule according to any one of claims 15-18.

10 49. A kit comprising one or more antibodies that bind to a polypeptide as recited in any one of claims 1-10; and a reagent useful for the detection of a binding reaction between said antibody and said polypeptide.

50. A transgenic or knockout non-human animal that has been transformed to express higher, lower or absent levels of a polypeptide according to any one of claims 1-10.

15 51. A method for screening for a compound effective to treat disease, by contacting a non-human transgenic animal according to claim 50 with a candidate compound and determining the effect of the compound on the disease of the animal.

ABSTRACT

This invention relates to a novel protein, termed CAA05410.2, herein identified as containing a Nuclear Hormone Receptor Ligand Binding Domain to functional equivalents of this protein and to the use of this protein, its functional equivalents and 5 nucleic acid sequence from the encoding genes in the diagnosis, prevention and treatment of disease.

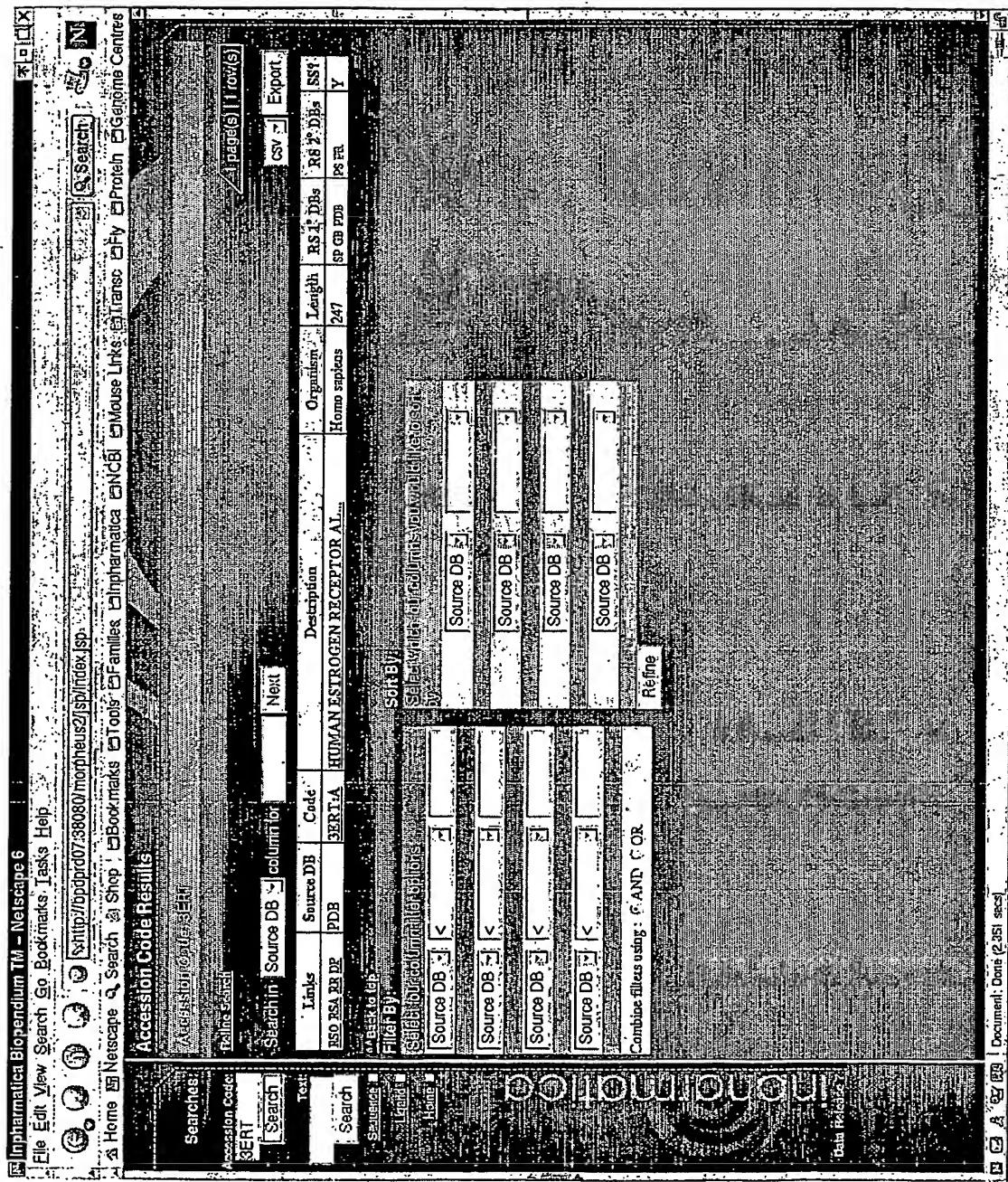


Figure 1

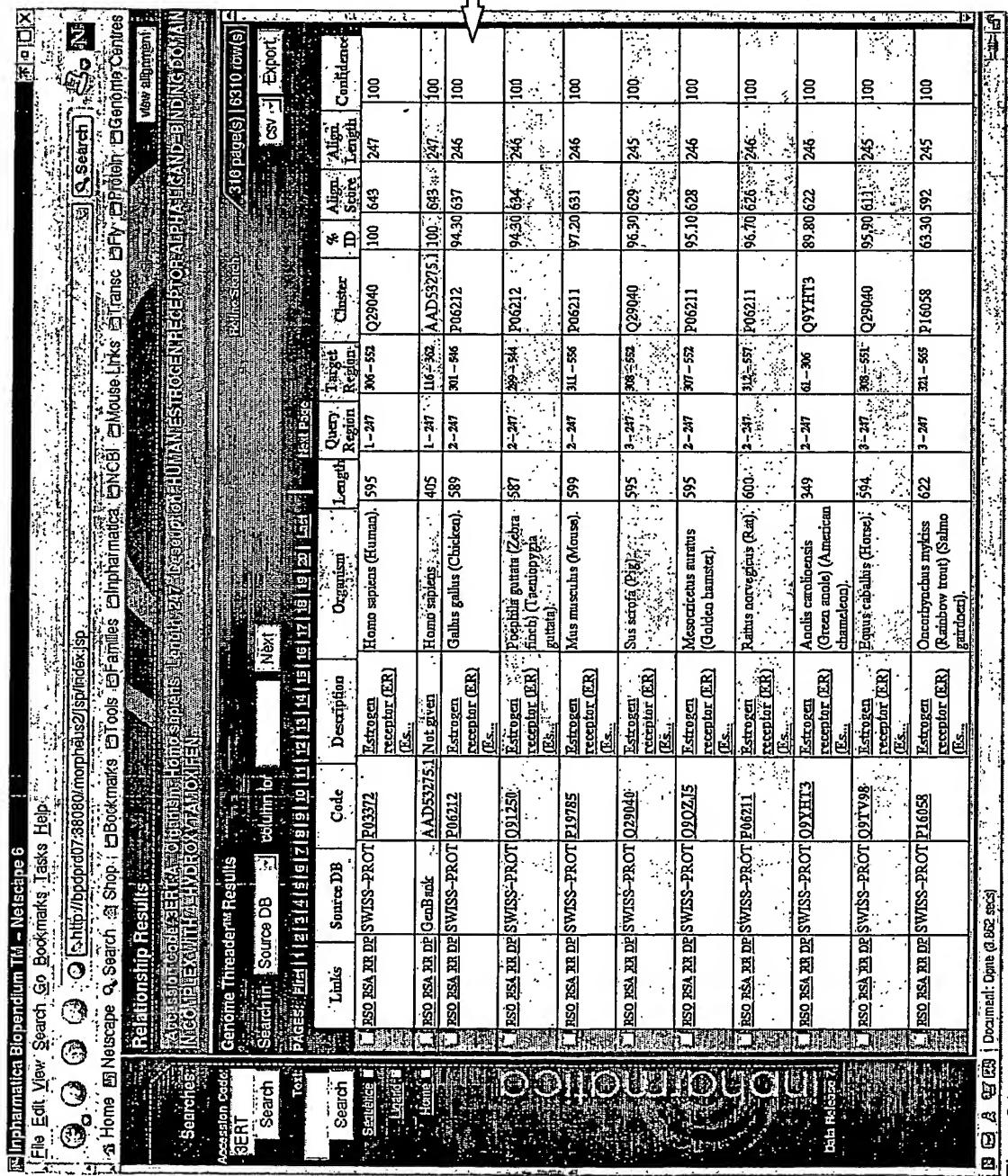


Figure 2A

Figure 2B

RMBI PSI-BLAST™ Results											
Relationship Results											
After 100 iterations, the following human proteins have been found to have significant similarity to the query protein:											
In order of decreasing similarity:											
55 pages(s) 1093 row(s)											
View alignment											
CSV Export											
Links	Source DB	Code	Description	Organism	Length	Query ID	Target Region	Cluster	% B61	E-value	
RBO	RBO	SWISS-PROT	Q22127	Nuclear hormone receptor, f.	Ceconobabtis elegans	289	3-107	181-205	Q22127	14 -4	2.0E-16
RBO	RBO	DP	CA192673	Not given	Ceconobabtis elegans	325	10-201	111-134	CA192673	10 -4	1.0E-15
RBO	RBO	DP	AA132663	minicircle cortical type I	Rattus norvegicus	100	9-52	47-100	AAB32663	224 -4	3.0E-15
RBO	RBO	DP	AA234981	lumbar receptor-3	Malacosoma disstria	227	1-112	81-105	AAB234981	13 -4	3.0E-15
RBO	RBO	DP	AAG15185	auditory receptor	Ceconobabtis elegans	273	41-244	63-273	AAQ15185	11 -4	4.0E-14
RBO	RBO	DP	AAK17971	lumbar receptor	Ceconobabtis elegans	493	21-146	321-459	AAK17972	11 -4	5.0E-12
RBO	RBO	DP	AAJ08044	Hypothetical protein	Ceconobabtis elegans	221	3-73	147-218	AAJ08043	19 -4	5.0E-11
RBO	RBO	DP	TK0812	lumbar receptor	Ceconobabtis elegans	474	9-124	33-146	AAK17972	11 -4	7.0E-10
RBO	RBO	DP	AA136900	NIR-2	Ceconobabtis elegans	232	8-109	116-223	AAB69001	11 -4	1.0E-7
RBO	RBO	DP	AA136900	Hypothetical protein	Ceconobabtis elegans	232	8-109	116-223	AAB69001	11 -4	1.0E-7
RBO	RBO	DP	AA166989	Not given	Ceconobabtis elegans	391	9-170	121-190	AA166989	17 -4	2.0E-7
RBO	RBO	DP	CA166989	Not given	Ceconobabtis elegans	252	3-132	49-251	CA166989	13 -5	9.0E-48
RBO	RBO	DP	CA1354410	Not given	Ceconobabtis elegans	217	3-237	1-217	CA1354410	19 -5	4.0E-47
RBO	RBO	DP	CA136098	Not given	Ceconobabtis elegans	900	10-247	161-411	CA136098	11 -5	4.0E-36
RBO	RBO	DP	AA175133	lumbar receptor	Ceconobabtis elegans	459	19-240	228-458	AA175133	18 -5	1.0E-30
RBO	RBO	DP	NIR-43	Not given	Ambystoma maculatum	416	4-97	236-311	AAB69281	11 -5	2.0E-9
RBO	RBO	DP	AA136928	developmental orphan	Expo...						
RBO	RBO	DP	CA103162	Not given	Homosapiens	614	4-70	341-518	CAA103162	11 -5	2.0E-8

Figure 2C

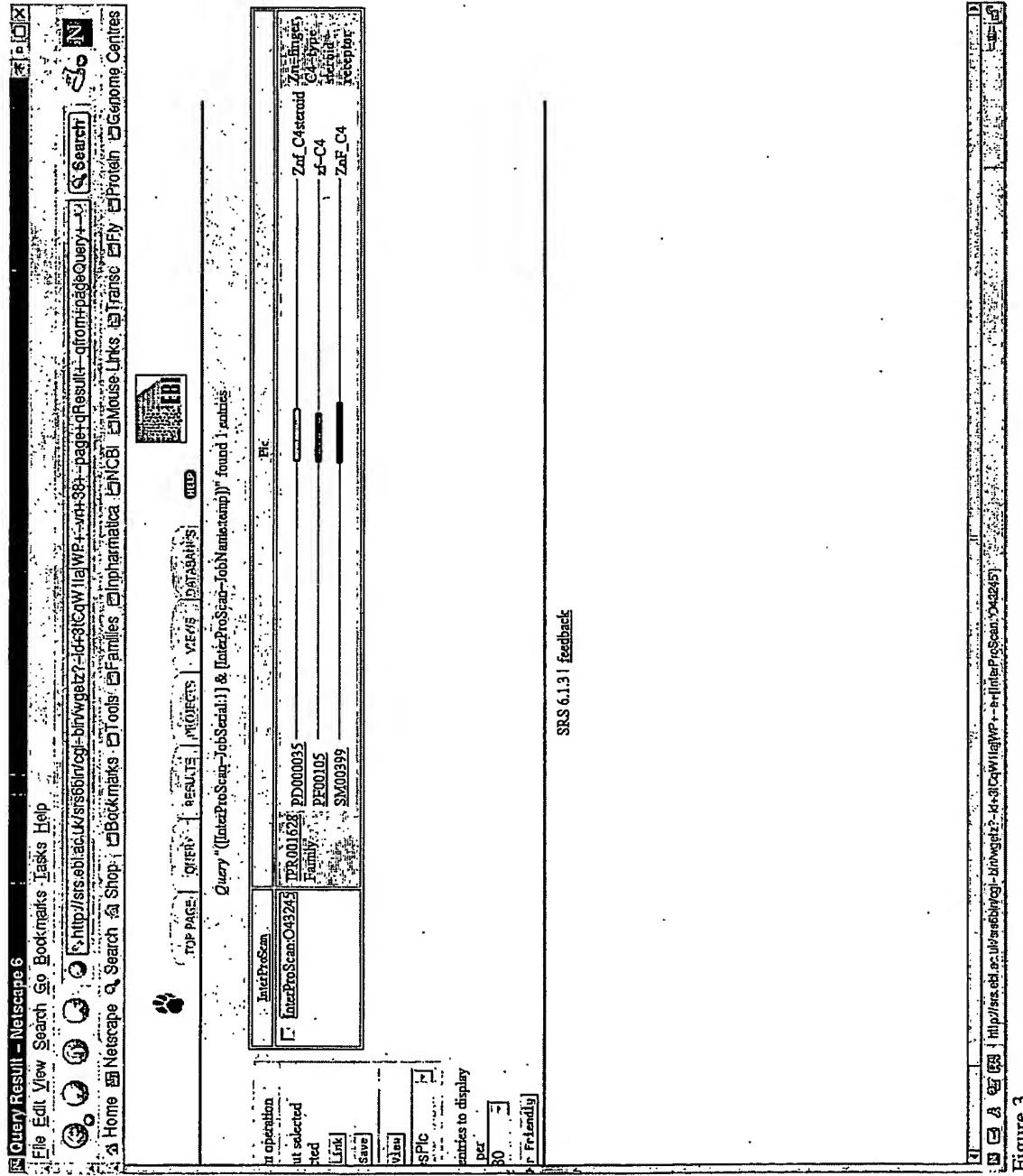


Figure 3

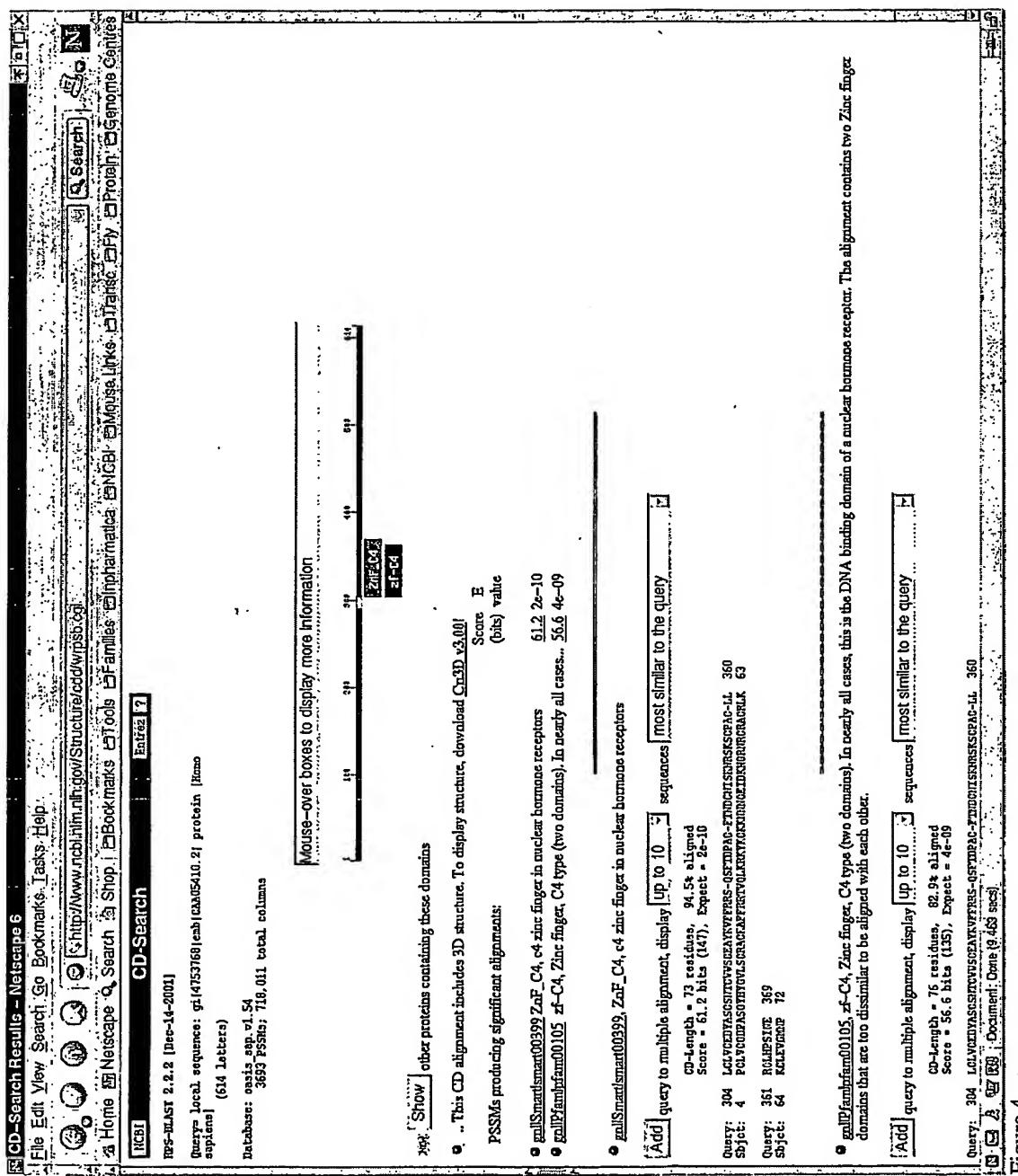


Figure 4

Figure 5A

Revised: October 24, 2001.

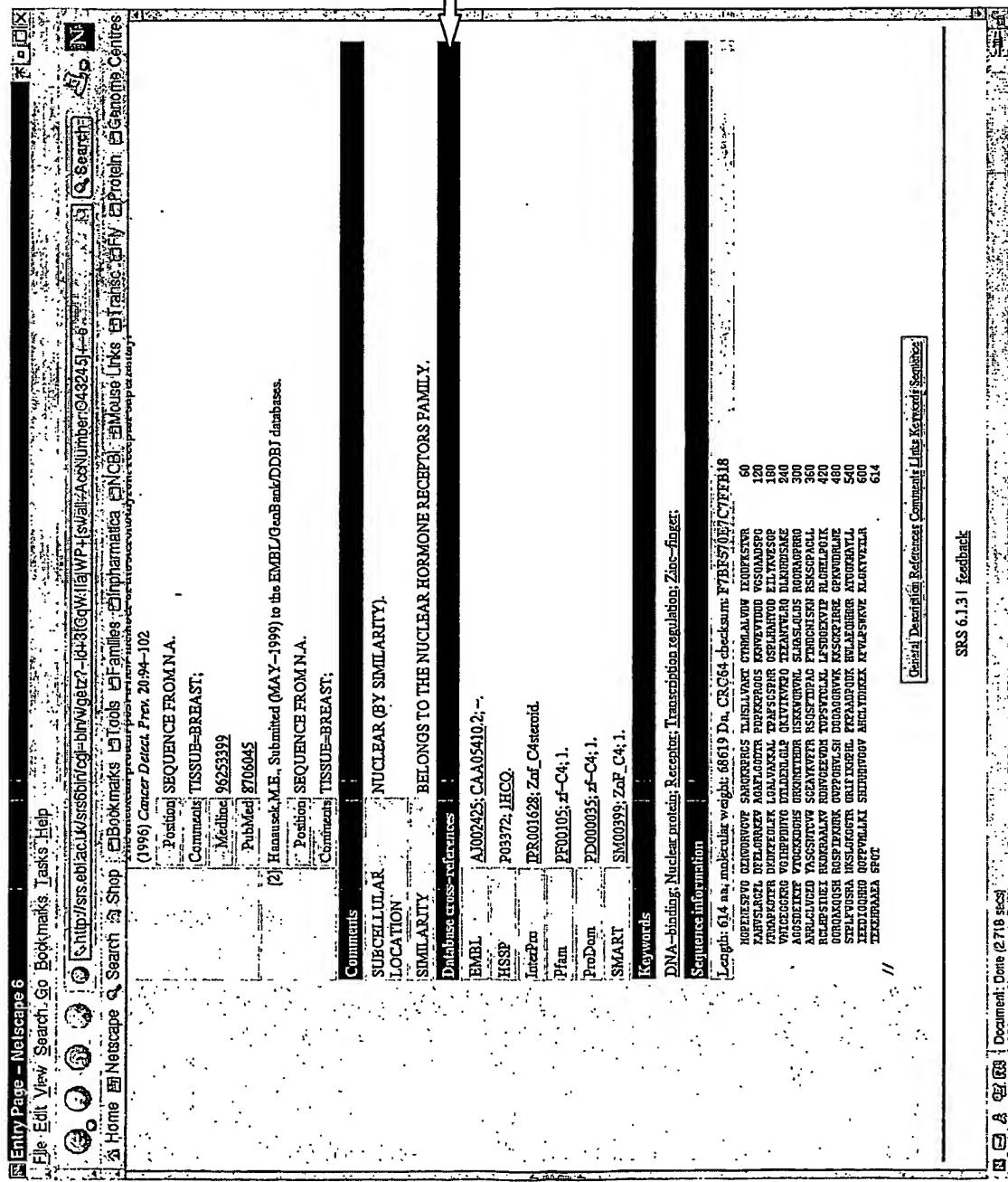


Figure 5B

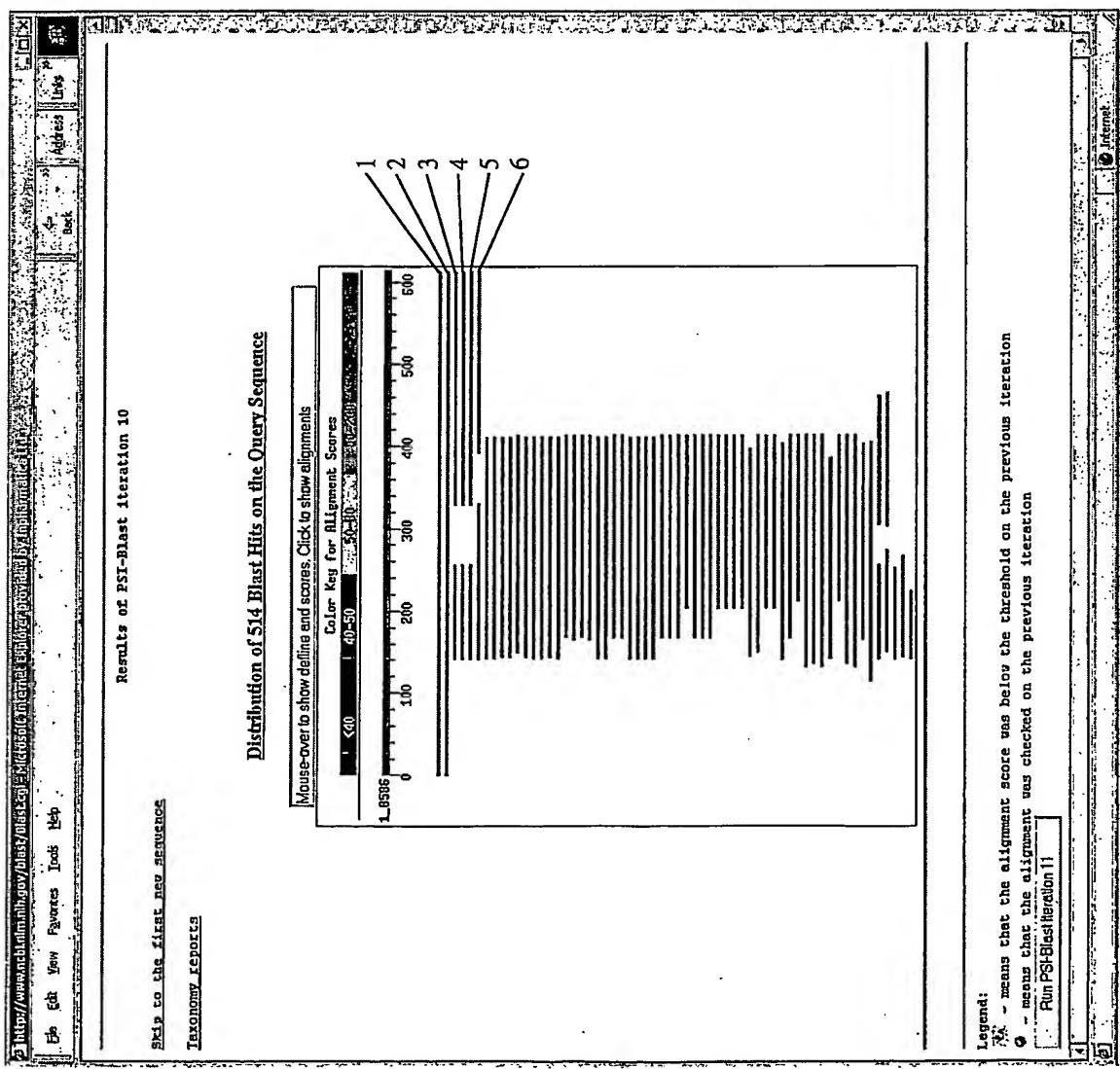


Figure 5C

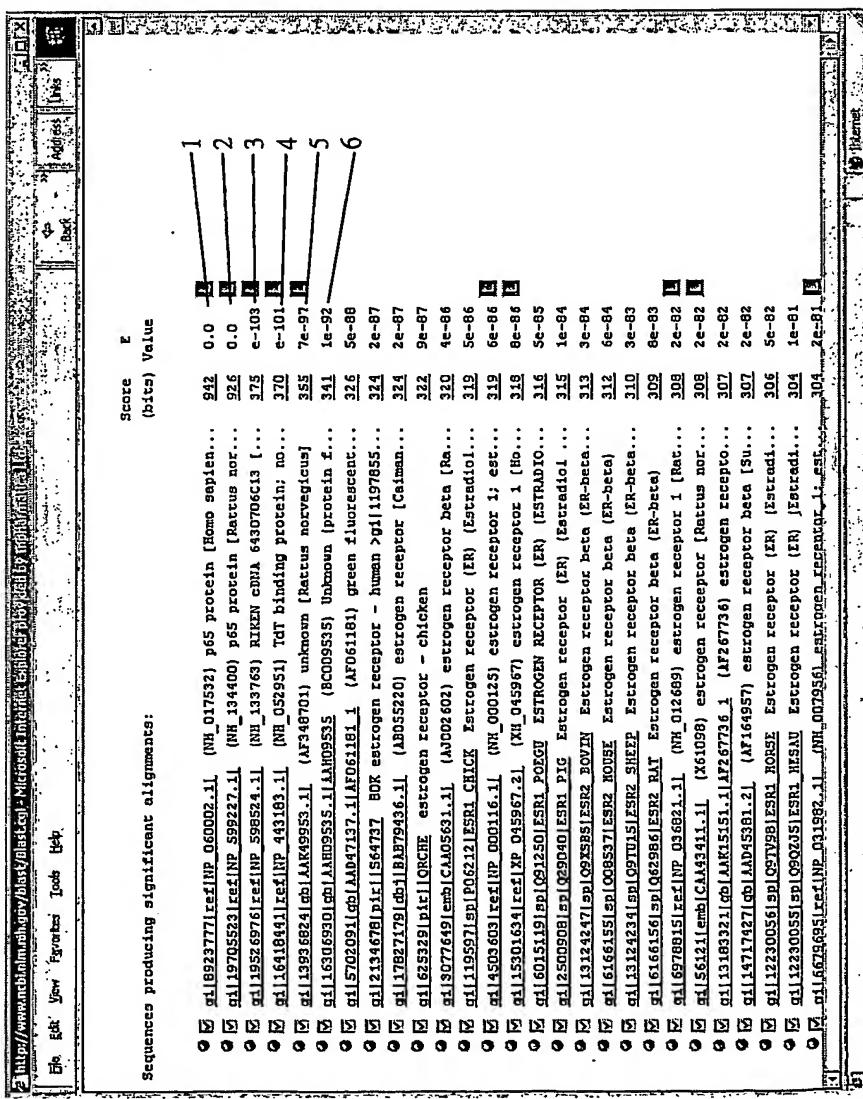


Figure 5D

Relationship Results										
Accession Code: CAA004102										
Search in: Source DB: PDB										
PAGES: First 1 12 13 14 15 16 17 18 Last Next Page										
Links	Source	Code	Description	Organism	Length	Query Region	Target Region	Cluster ID	Align. Length	Confidence
ESO-ESR-DE-PDB	SERTA	HUMAN ESTROGEN RECEPTOR AL...	Homosapiens	247	394 - 684	15 - 247	1QKU	14,60	256	239
ESO-ESR-DE-PDB	SERDB	HUMAN ESTROGEN RECEPTOR AL...	Homosapiens	238	394 - 603	16 - 238	1QKU	14,90	252	228
ESO-ESR-DE-PDB	SERDA	HUMAN ESTROGEN RECEPTOR AL...	Homosapiens	246	394 - 602	16 - 246	1QKU	14,80	251	237
ESO-ESR-DE-PDB	1OKUA	WILD TYPE ESTROGEN NUCLEAR...	Homosapiens	250	394 - 602	20 - 250	1QKU	14,80	251	237
ESO-ESR-DE-PDB	1OKUB	WILD TYPE ESTROGEN NUCLEAR...	Homosapiens	247	394 - 602	17 - 247	1QKU	14,80	251	237
ESO-ESR-DE-PDB	1OKUC	WILD TYPE ESTROGEN NUCLEAR...	Homosapiens	247	394 - 602	17 - 247	1OKU	14,80	251	237
ESO-ESR-DE-PDB	1OKTD	MUTANT ESTROGEN NUCLEAR RE...	Homosapiens	248	394 - 603	17 - 248	1OKU	14,70	247	238
ESO-ESR-DE-PDB	1TEREA	HUMAN ESTROGEN RECEPTOR L1...	Homosapiens	235	403 - 600	19 - 235	1QKU	14,30	243	223
ESO-ESR-DE-PDB	1TEREB	HUMAN ESTROGEN RECEPTOR L1...	Homosapiens	235	403 - 600	19 - 235	1OKU	14,30	243	223
ESO-ESR-DE-PDB	1TEREC	HUMAN ESTROGEN RECEPTOR L1...	Homosapiens	235	403 - 600	19 - 235	1OKU	14,30	243	223
ESO-ESR-DE-PDB	1TERED	HUMAN ESTROGEN RECEPTOR L1...	Homosapiens	235	403 - 600	19 - 235	1QKU	14,30	243	223
ESO-ESR-DE-PDB	1TEREE	HUMAN ESTROGEN RECEPTOR L1...	Homosapiens	235	403 - 600	19 - 235	1QKU	14,30	243	223
ESO-ESR-DE-PDB	1TEREF	HUMAN ESTROGEN RECEPTOR L1...	Homosapiens	235	403 - 600	19 - 235	1OKU	14,30	243	223
ESO-ESR-DE-PDB	1AKT	HOMOLOGOUS-EXTENSION-BASED...	N/A	237	394 - 599	10 - 237	1QKU	14,50	236	234
ESO-ESR-DE-PDB	1OKMA	HUMAN OESTROGEN RECEPTOR B...	Homosapiens	230	395 - 598	8 - 228	1QKM	13,50	236	229
ESO-ESR-DE-PDB	1IMPA	THE 21 ANGSTROM RESOLUTION...	N/A	232	403 - 598	10 - 231	1LBD	10,90	226	137

Figure 6

Relationship Results											
Searches:		Source DB			Description			Protein			
Links	Source DB	Code	Code	Code	Target	Length	Region	Target	Length	Region	
ESO-ESA-DE-GenBank	AAA517251	Androgen receptor	Hom sapiens		730	162-485	227-537	AAA517351	31	2	6.0E-31
ESO-ESA-DE-GenBank	AAA517711	Androgen receptor	Hom sapiens		917	261-578	538-765	P10215	31	2	6.0E-31
ESO-ESA-DE-GenBank	AAA517751	Androgen receptor	Hom sapiens		918	261-578	538-787	P10275	31	2	6.0E-31
ESO-ESA-DE-GenBank	A1A4e59211	Androgen receptor	Hom sapiens		906	129-578	410-775	P10275	31	2	6.0E-31
ESO-ESA-DE-SWISS-PROT	P08235	Minerolocorticoid receptor	Hom sapiens (Human)		984	67-311	367-632	P08235	34	2	8.0E-31
ESO-ESA-DE-SWISS-PROT	Q19952	IGF1 receptor	Oryctolagus cuniculus (Rabbit)		475	216-412	524-531	BAAS81001	24	2	8.0E-31
ESO-ESA-DE-GenBank	CAB61281	Interferon receptor	Crocidolylus milbicus		933	241-428	6-169	CAB612631	24	2	8.0E-31
ESO-ESA-DE-SWISS-PROT	Q97776	Androgen receptor	Bulinus fibris collinus (Collared brown leech)		884	192-578	365-793	Q97776	31	2	1.0E-30
ESO-ESA-DE-SWISS-PROT	P49843	Glucocorticoid receptor (G...)	Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri)		758	171-577	281-631	P49843	34	2	1.0E-30
ESO-ESA-DE-GenBank	AAAC52431	Mediator orphan receptor H2...	Rattus norvegicus		598	166-599	146-596	AAC21431	24	2	1.0E-30
ESO-ESA-DE-GenBank	AAA621101	IPAR gamma	Mus musculus		475	216-428	1-243	BAA981001	24	2	1.0E-30
ESO-ESA-DE-SWISS-PROT	Q73673	Chloroenthol	Parathrybys diffriculus (Phunder)		807	188-595	331-696	Q74673	33	2	2.0E-30
ESO-ESA-DE-GenBank	AAAG423621	Progesterone receptor	Xenopus laevis		732	83-377	128-600	AAAG423621	28	2	2.0E-30
ESO-ESA-DE-GenBank	CAB612631	Progestins receptor	Trichomyctes elegans		393	224-423	6-169	CAB612631	24	2	2.0E-30
ESO-ESA-DE-SWISS-PROT	P49699	Androgen receptor	Oryctolagus cuniculus (Rabbit)		709	32-578	105-578	P49699	30	2	3.0E-30
ESO-ESA-DE-SWISS-PROT	Q29131	Mineralocorticoid receptor...	Tupai gl. belangeri (Common tree shrew)		977	67-671	365-945	Q29131	34	2	3.0E-30
ESO-ESA-DE-SWISS-PROT	P49700	Ultrashortad protein	Bombyx mori (Silk moth)		462	248-396	23-205	P49700	22	2	3.0E-30
ESO-ESA-DE-SWISS-PROT	P19091	Androgen receptor	Mus musculus (Mouse)		819	210-578	581-768	P15207	30	2	4.0E-30

Figure 7A

Figure 7B

CAA05410 . 2 3ERT	MGGENESPVGEMGRVCVFSARQKRPRCSTLNSLLVARTCYNMLALYDWLEQDFKSTWR -----
CAA05410 . 2 3ERT	KAHVSLRGFLDFELGRKEVAGAFLGGDTRPDPKPRGGSKKNVEVYDDVGSQAADSPG 120 -----
CAA05410 . 2 3ERT	KMAPKGTFRDKDKEGLFKLGAVALAKKALITDAFSCSPNRSPLHAFHYGDELLYKVESGP 180 -----
CAA05410 . 2 3ERT	VNICEGGKRGKGVGVIHPPDNYGDTLDENLGLPQKIVIKVKPQTEEANTWLRQDLKNHNSAKE 240 -----
CAA05410 . 2 3ERT	AGGSDEI1KTFVTGCKKDGHSGRKNMTHDRNSKKWQRVNNSLMASLQOLDSRGGAGP芮RG 300 -----
CAA05410 . 2 3ERT	APRLCLVCEDYASCSNTICVWSCEAYKVFRRRSQSFTDPACFTNDCNISKNRSKSCPACLL 360 -----
CAA05410 . 2 3ERT	RCLHPSINEIRKDRAALKVDRDNVGEEVDMTGPSWTCLKLFSDEGEKVIPRLGHELPGIK 420 -----LALSITADQMVSLALDAEPPILYSEYDPTPFESEASMMGLL 346 -----
CAA05410 . 2 3ERT	GGRQAKQSHRGSP1PKNPKRGWPPGHVLSNDGGAGGRVWKKSKCKPIRREGPKWDRLN- 479 TMNADRELVH-MINWAKRVPGFVDLTLHDQVHLLECAWLEIIMIGLYWRSMEHPGKLLFA 405 -----
CAA05410 . 2 3ERT	-----ESTPLFWGSRANKSLGKGTRGRIFIKPHLFEAADPDK-- 520 PNLLLDRNQGKCVEGMVEIDMLLATSS--RFRMMNLQGEEFVCLKSILLNSGVYTFLS 463 -----
CAA05410 . 2 3ERT	---HMLAEQHHMRATGGK---MAYLLIEEDIG-OHFGQGFPVMTL[KI]SHIRHMVGGYAH 572 STLKSLEEKDHIIHVLVDKITDTLHILMAKAGLT[QQHQH]RQLI[L]LSHIRHMSNKGM 523 -----***-*-*-*-*DIMER HELIX***-*-* -----
CAA05410 . 2 3ERT	C[Y]DMKEKEFVLP5WVKEKLGKYVETLRTKEHRAAEASPOT 614 HLYSMKCKNVVP---LYDLLEMDAHLHAP----- 552 -----***-*-*-* -----

Figure 8: Genome Threader alignment (3ERT numbering corresponds to position in the full-length sequence rather than crystallised domain)

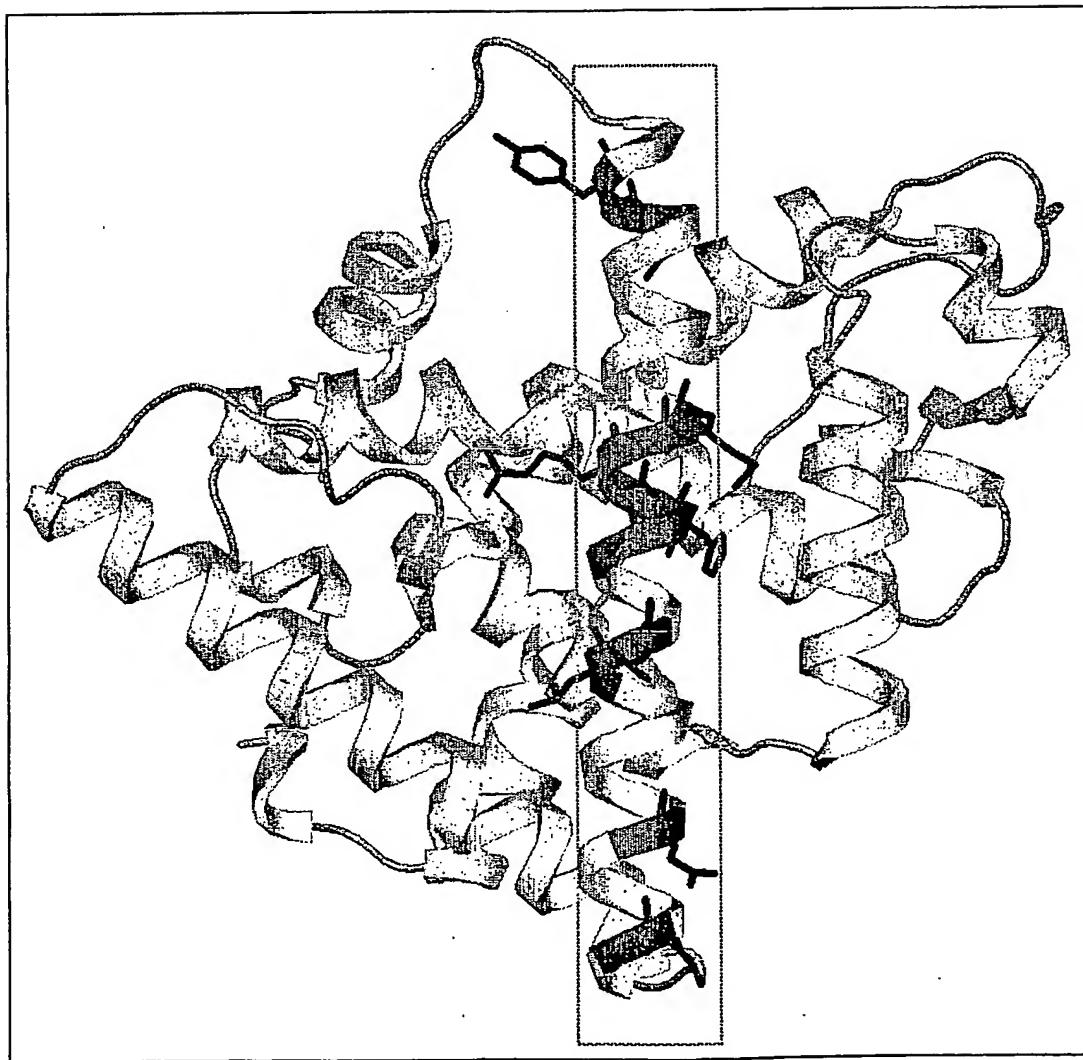


Figure 9: Dimerisation helix is enclosed in dotted box.

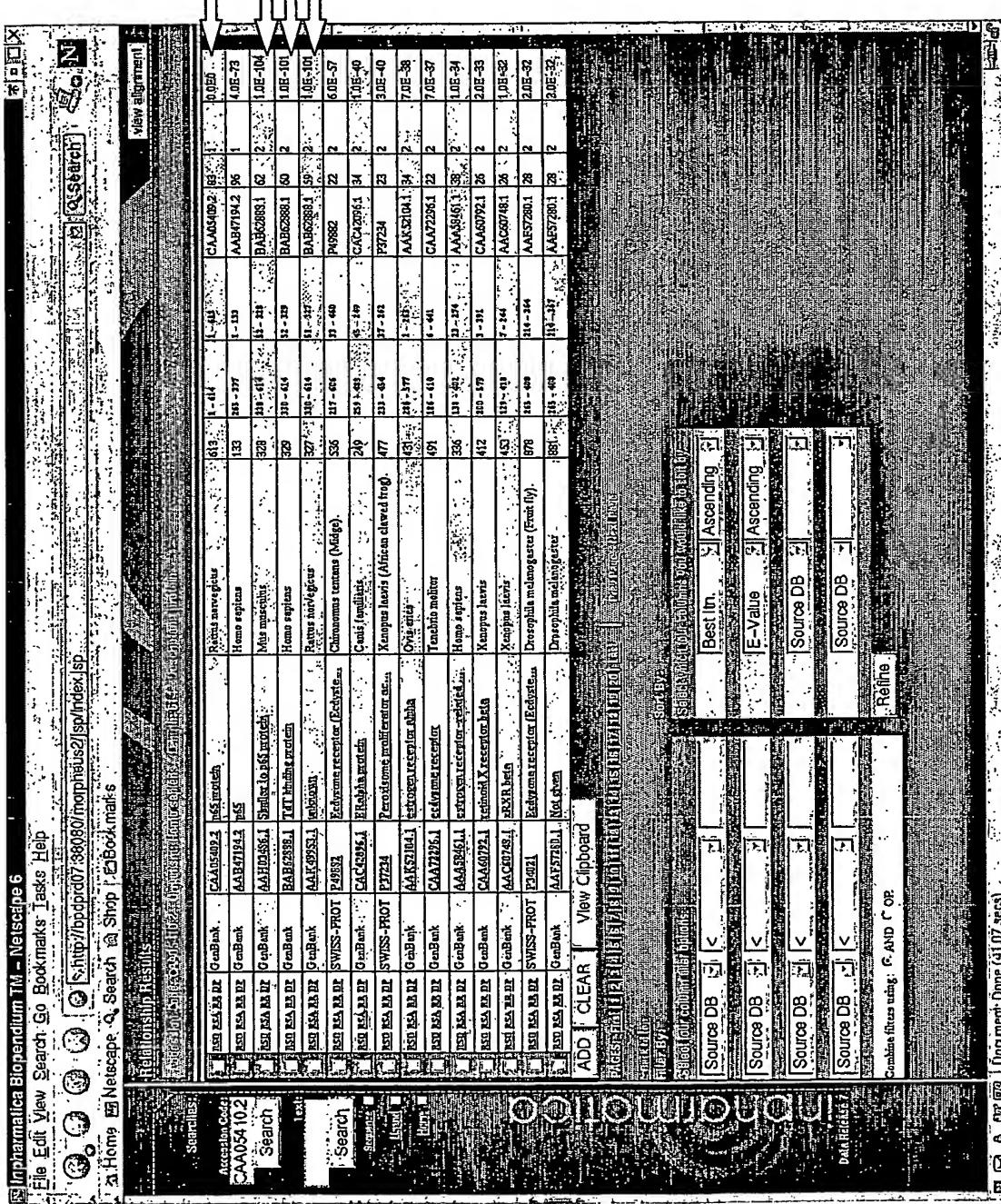


Figure 10

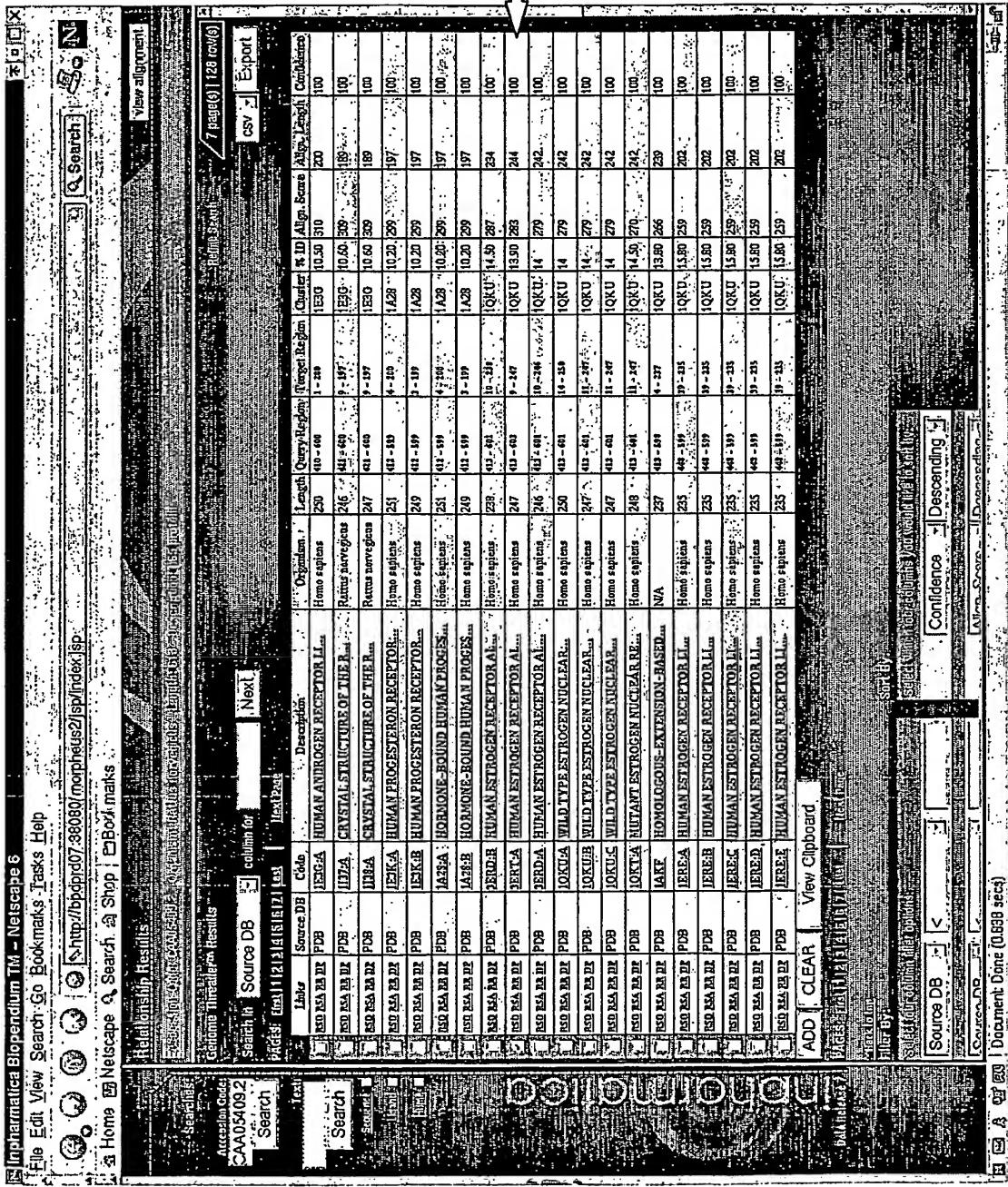


Figure 11

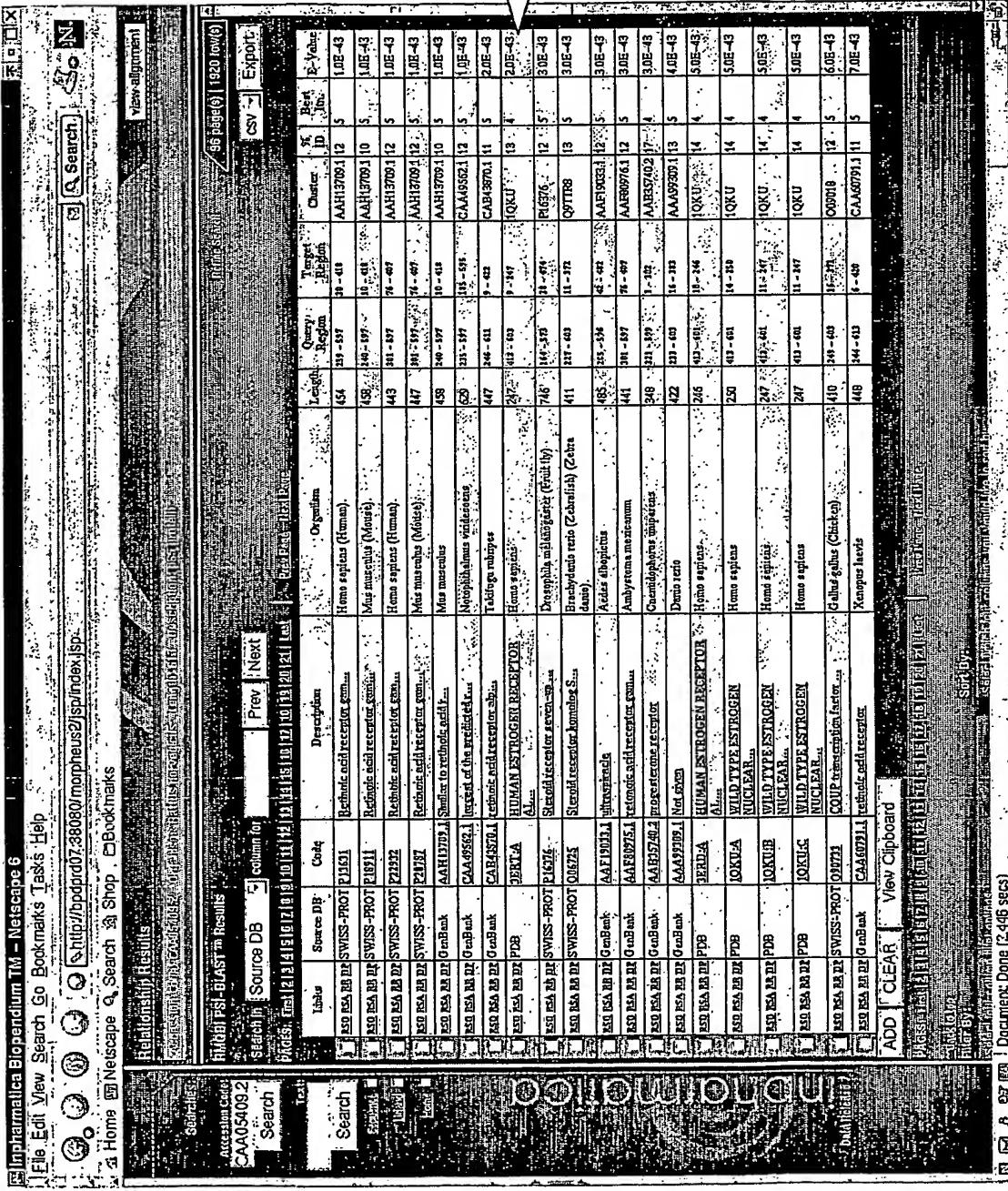


Figure 12

3ERT	CAA05409 .2	MDPENVRAGGDMREKVRLSSARQ-RLRCSTKNSLLVARTC--VILRLVWDWIEQDLKSTCE	57
	CAA05410 .2	MGPNESPVGGMWGRVCVFSSARQKRPRCSTLNSSLVARTCYNMLALVWDWIEQDFEKSTWR	60
3ERT	CAA05409 .2	KAHVSLLRGYLDQQLQCGRRKEVAGA- LGGDTRPDPKKPRGGSKKNVVEVYGDDVGSQAADS	116
	CAA05410 .2	KAHVSLLRGELDFEL- -GRRKEVAGAFLGGDTRPDPKKPRGGSKKNVVEVYDGVGSQAADS	118
3ERT	CAA05409 .2	PRKQLAAKGTFRDKDKEIAFLFKLGEVAKKALSSAITWFPNSVSPLHAHYGDEILYKDES	176
	CAA05410 .2	PGKRMAPKGTFRDKDKEFLFKLGEVAKKALTPAFSCSPNRSPLHAHYGDEILYKVES	178
3ERT	CAA05409 .2	GLVNISEGGKRGVEIHPDNFGITLDEDLGFPQIIVINVKPQTEEANTWYRQDLKYHNS	236
	CAA05410 .2	GPVNICEGGKRGVGIIHPPDNYG-DTLDENIGLPQKIVIKVVKPQTEEANTWLRQDLKNHNS	237
3ERT	CAA05409 .2	ANEAGYSDENKTFVRGCRKDGHSERNNMTTGDRNSKKAQPYNESSLMASLALDSRGKAAGP	296
	CAA05410 .2	AKEAGGSDEIKTFTVTCGCKKDGHSGRKNMTHDRNSKKWQRVNLSLMASLQDLSRGGRAGP	297
3ERT	CAA05409 .2	RRGARRLCLVCEDYASCSNTCVWSCEAYKVFERRSQSFTIDPACFTNDCNISKNRSKSCPA	356
	CAA05410 .2	RRGARRLCLVCEDYASCSNTCVWSCEAYKVFRRSQSFTIDPACFTNDCNISKNRSKSCPA	357
3ERT	CAA05409 .2	-LAISLTADQMVSSLDAEPPILYSEYDPTTRPFSSEASMM	343
	CAA05410 .2	CLLRCLQPSINEIRKDKRAALNVRDNVGEEVDMTGPSWTCLKLLFSDGEKVIPRLAHELP	416
3ERT	CAA05409 .2	CLLRCLQPSINEIRKDKRAALKVARDNVGEEVDMTGPSWTCLKLLFSDGEKVIPRLGHELP	417

Figure 13 Part I

3ERT	GLITNLADRELVH-MINWAKRVPGFVVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKL	402
CAA05409.2	GIKRGQRQAAEEESHRGSPIPKRRKGWPPGHVLSNDGGAGGRVWKKSCCKPIRREGPKWNDAR	476
CAA05410.2	GIKGGGRQAKQQSHRGSPIPKRNKGWPPGHVLSNDGGAGGRVWKKSCCKPIRREGPKWNDAR	477
3ERT	LFAPNLLLDRNQGKCVEGMVEIFDMLLATSS--RERMMNLIQGEEEFVCLKSIIILNSGVYT	460
CAA05409.2	LN-----ESTTIVLGSRANKALGKGGTRGRYIKHPHILEKYAADPQD	518
CAA05410.2	LN-----ESTPLFWGSRANKSLIGKGGTRGRIFIKHPHLFKFAADPQD	519
3ERT	FLSSTLKSLEEKDHILHRVLDKITDTLHILMAKAGLTLQQQHQORLAQOLLILSHIRHMSNK	520
CAA05409.2	K-----HMLAEQLHM/RATGGK-----MAYLLIEDDIE- QH HGQFAVFLIKISHIRHM/VEG	568
CAA05410.2	K-----HMLAEQHHM/RATGGK-----MAYLLIEDDIG- QH HGQGPVMLIKISHIRHM/VGG	569
	*****DIMER HELIX*****	
3ERT	GMEHLYSMKCKNVVP---LYDILLEMIDAHRLHAP-----	552
CAA05409.2	VAHCLYDMKVIQFVLPSWKVEKLRYVETLRTENEHRAAEASPQT	613
CAA05410.2	VAHCLYDMKEKKFVILPSWKVEKLGYVETLRTKEHRAAEASPQT	614

Figure 13 Part II

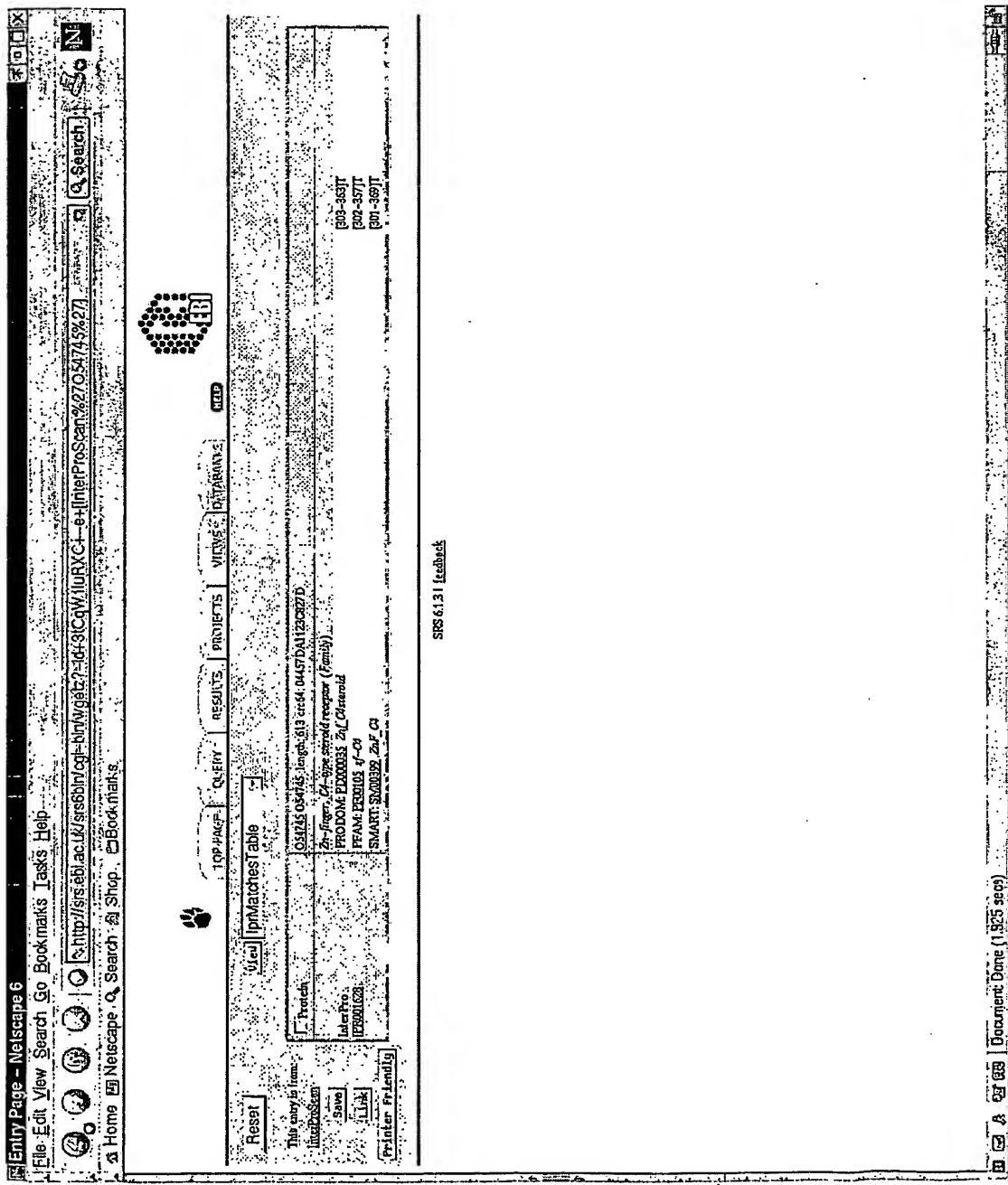


Figure 14

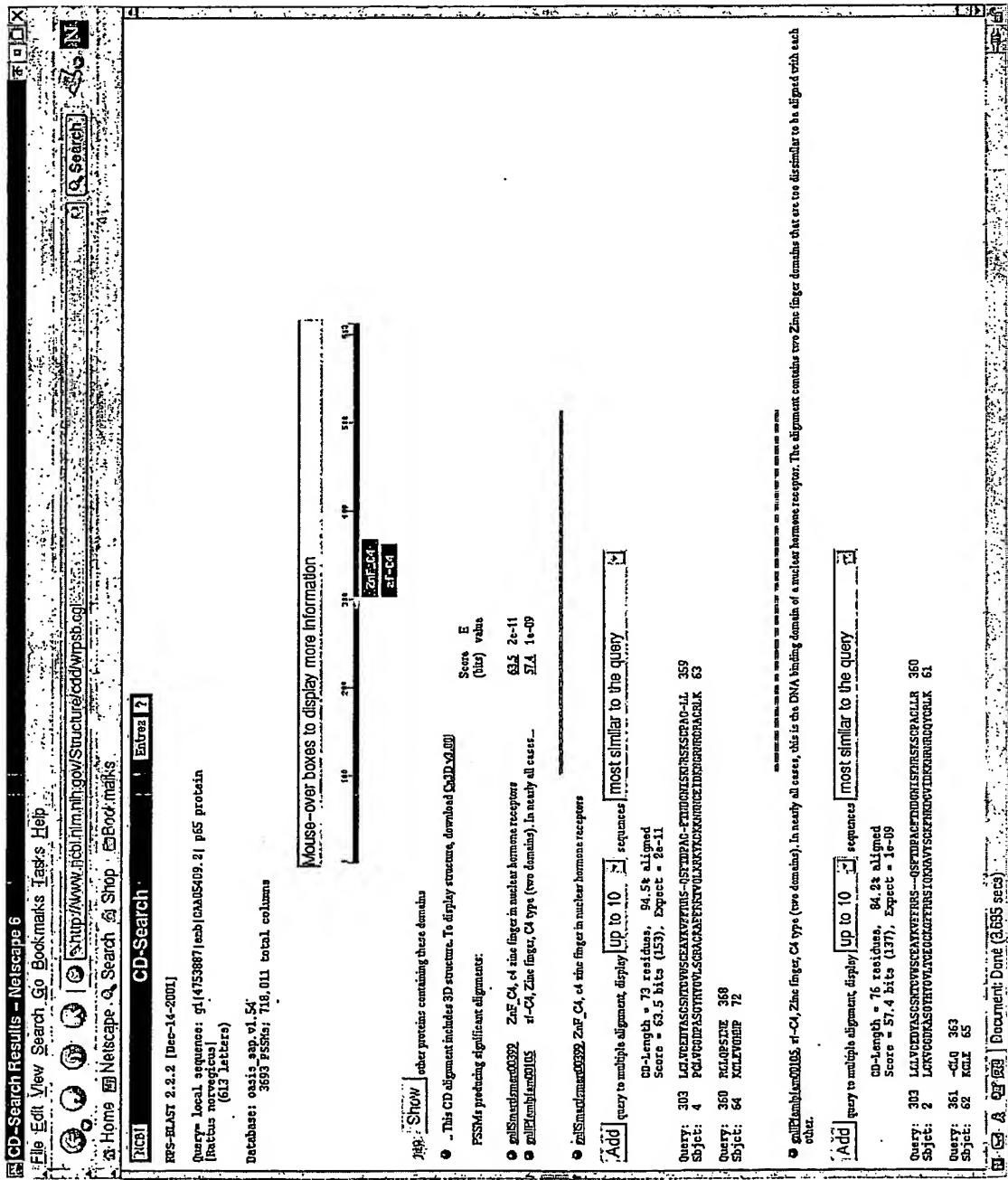


Figure 15

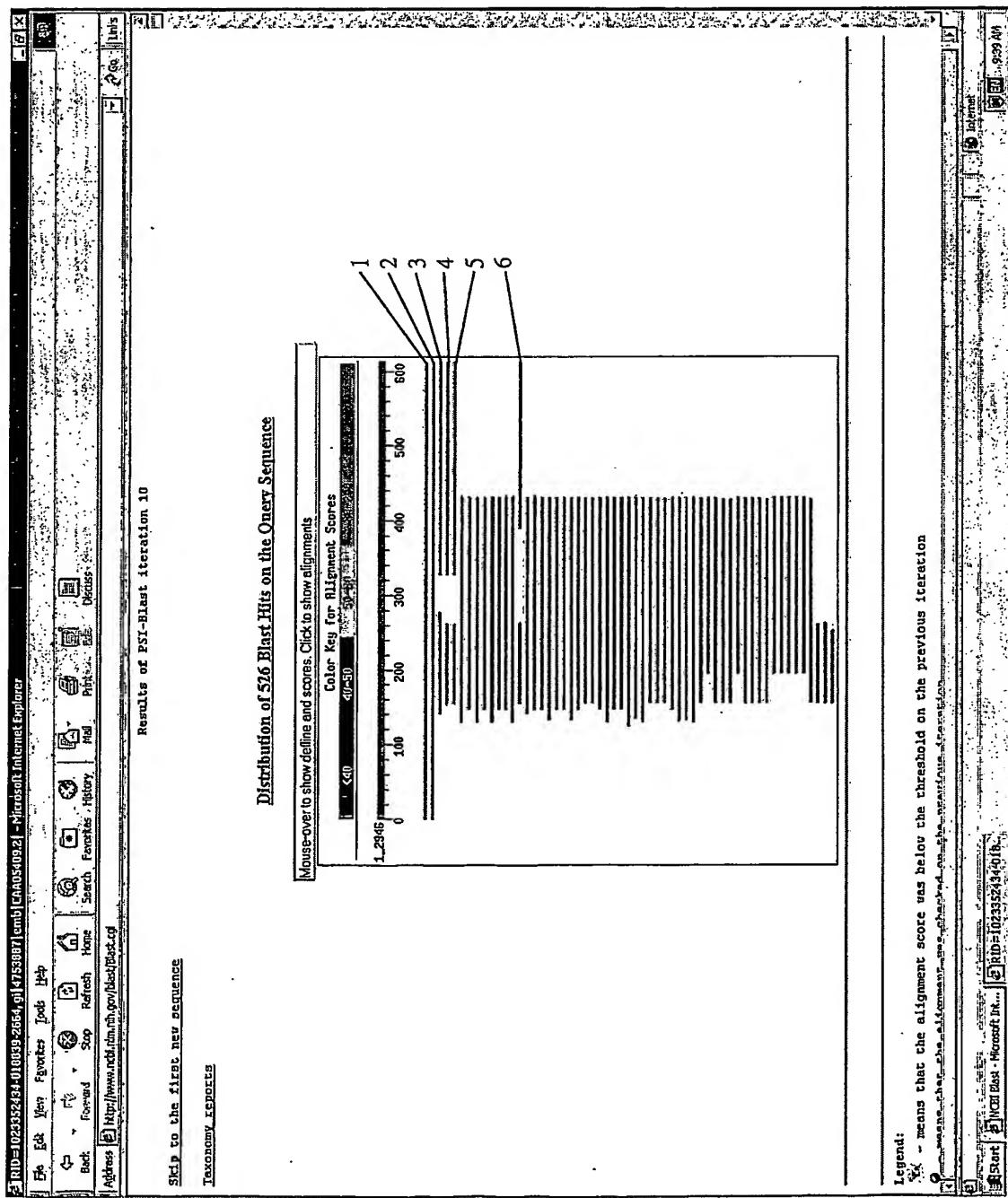


Figure 16A

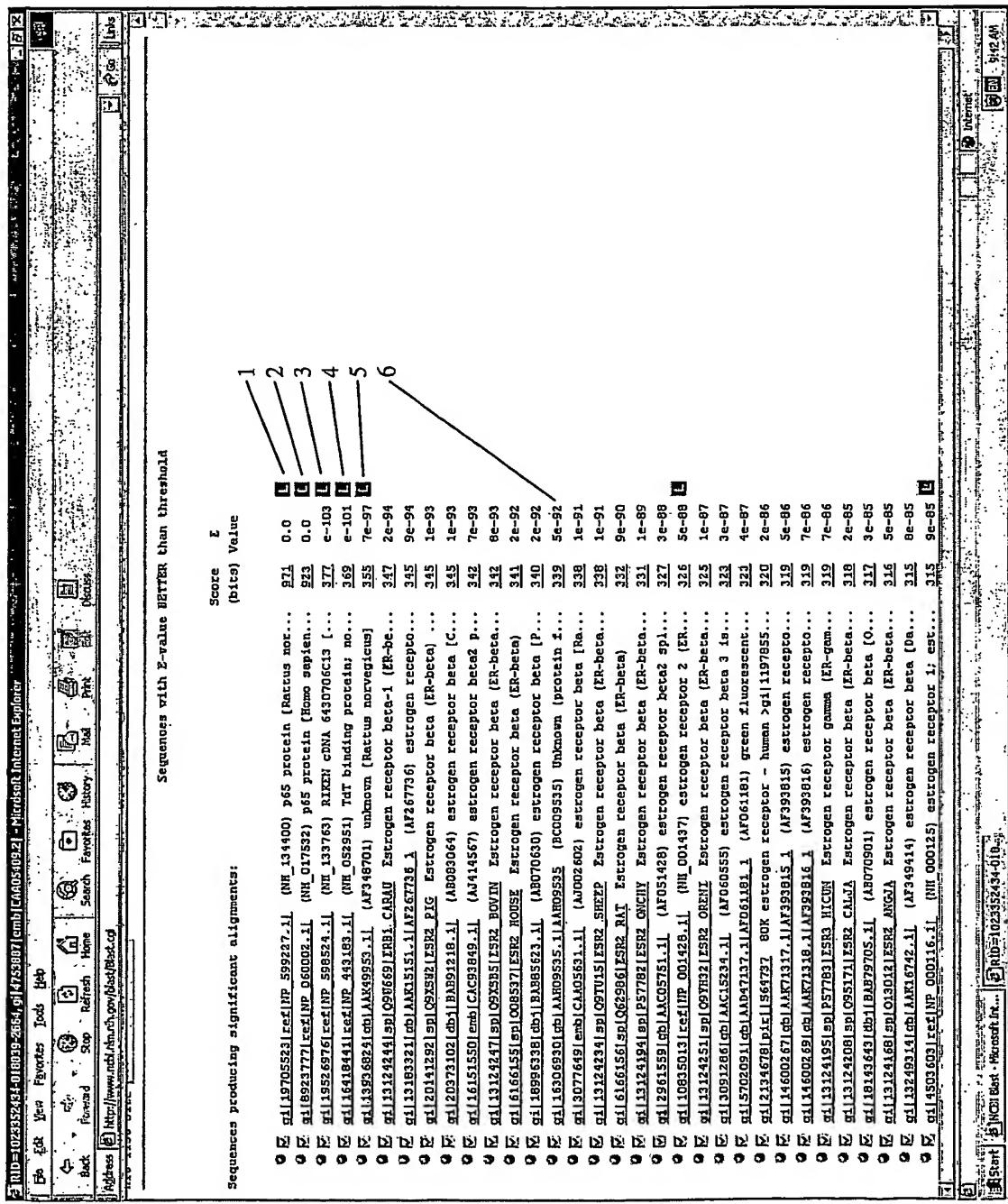


Figure 16B

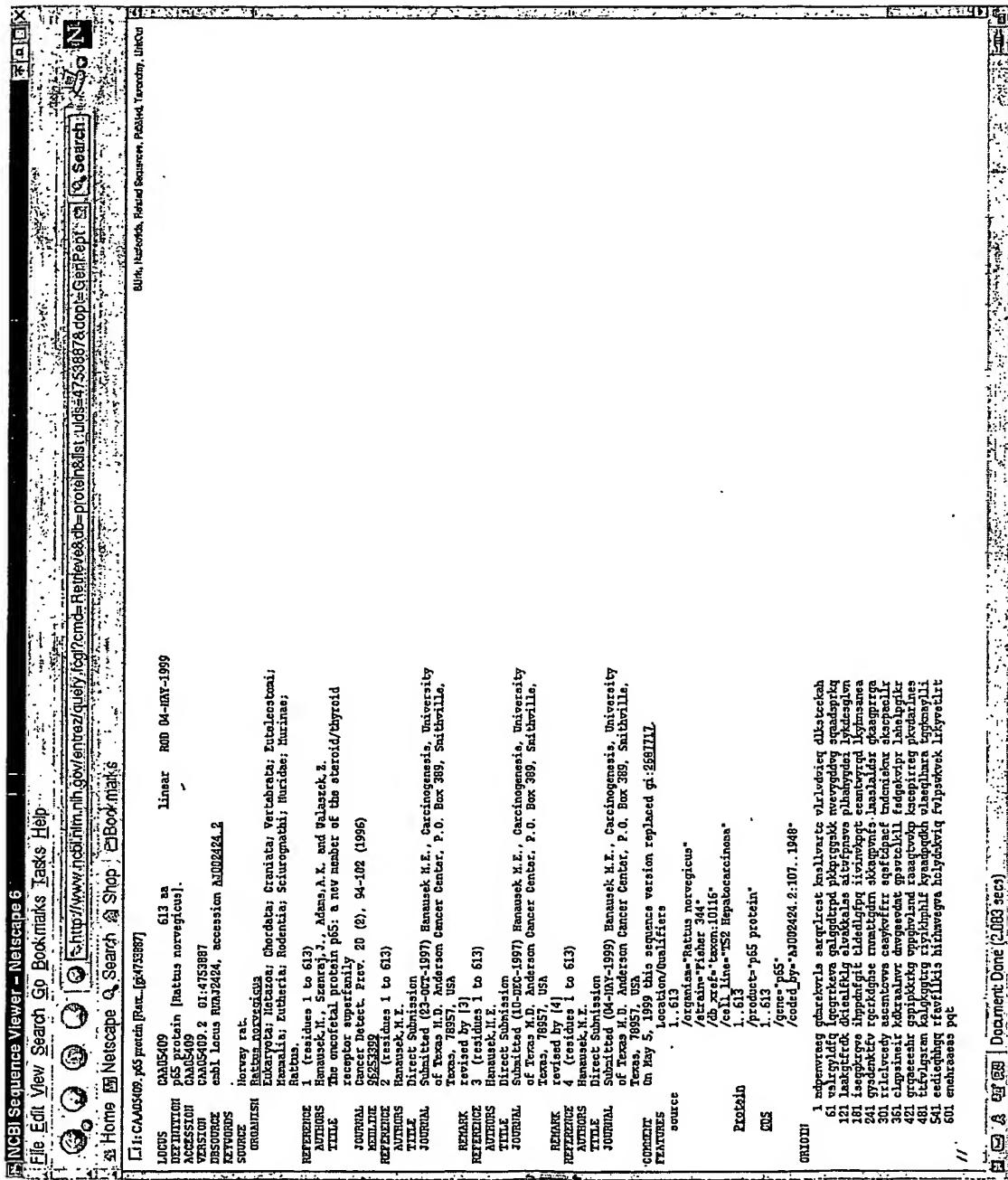


Figure 17

3ERT:A

1 (306) 15 (320)

247 (552)

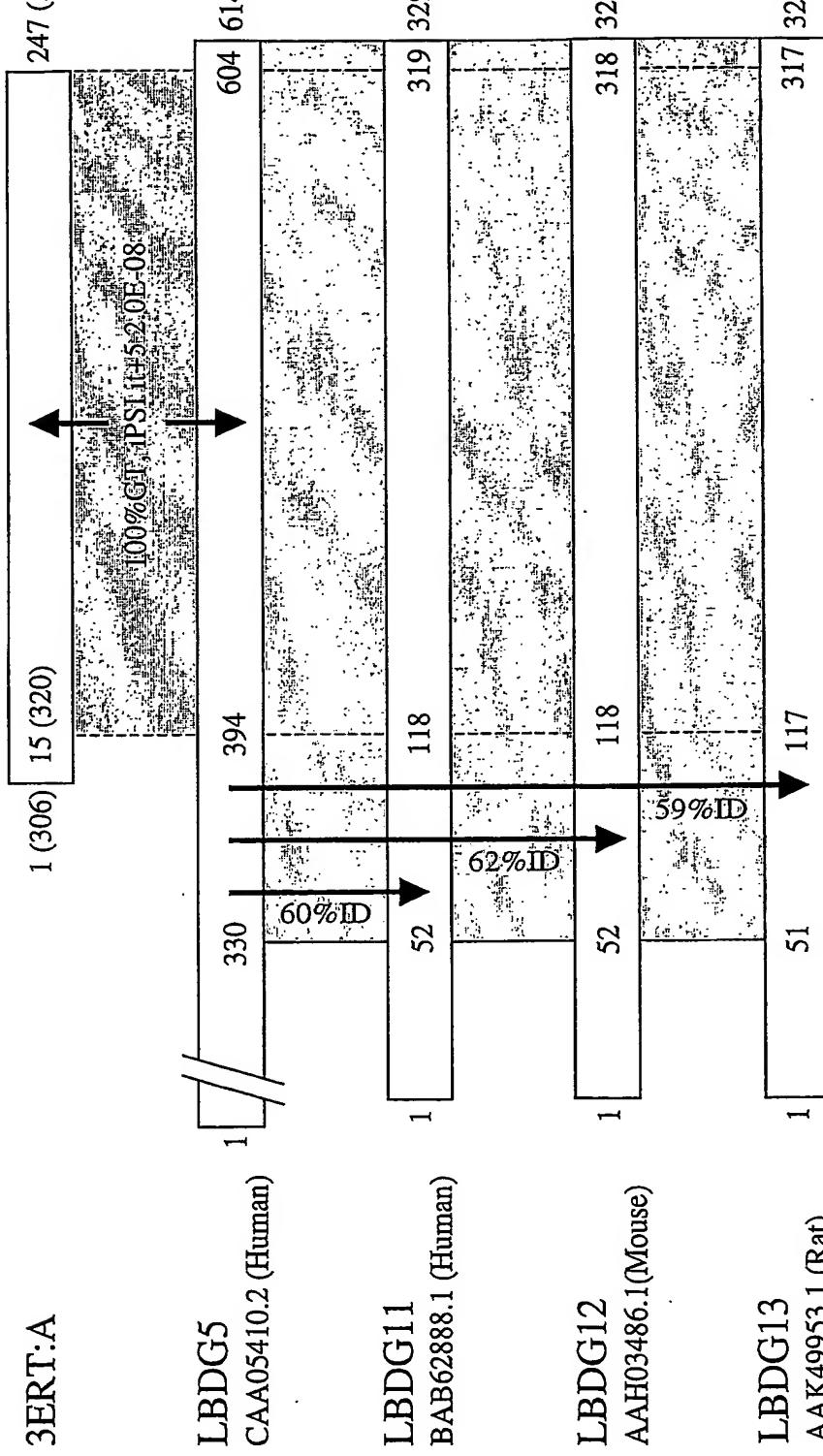


Figure 18: relationships between 3ERT:A, CAA05410.2 (LBDG5), BAB62888.1 (LBDG11), AAH03486.1 (LBDG12) and AAK49953.1 (LBDG13).

Figure 19 Part I

3ERT		MIGLVMRSMEEHPG-KLLFAPNLLLDRNQGKCVEGMVEIFDMLLATSS--RFRMMNLQGEE	444
CAA05410.2	(LBDG5)	SCKPIRREGPKWW-DRLN-----ESTPLFWGSRANKSLGKGGTRGRI	503
BAB62888.1	(LBDG11)	SCEPIRREGPKWDPARLN-----ESTTFVLGSRANKALGMGGTRGRI	232
AAH03486.1	(LBDG12)	SCEPIRREGPKWDPARLN-----ESTTFVLGSRANKALGMGGTRGRI	231
AAK49953.1	(LBDG13)	SCEPIRREGPKWDPARLN-----ESTTFVLGSRANKALGMGGTRGRI	230
3ERT		FVCLKSIIILLNSGVYTFLSSTLKSLEEKDHIIHVLVDKITDTLILHMAKAGLTQQQHQQL	504
CAA05410.2	(LBDG5)	FIKHPHFLFKFAADPQDK-----HWLAEQHHMRATGGK-----MAYLLIEEDIG-QHHGQGF	553
BAB62888.1	(LBDG11)	YIKHPHFLFKYAADPQDK-----HWLAEQHHMRATGGK-----MAYLLIEEDIR-D-----	276
AAH03486.1	(LBDG12)	YIKHPHFLFKYAADPQDK-----HWLAEQHHMRATGGK-----MAYLLIEEDIR-D-----	275
AAK49953.1	(LBDG13)	YIKHPHFLFKYAADPQDK-----HWLAEQHHMRATGGK-----MAYLLIEEDIR-D-----	274
3ERT		AQLLILISHIRHMSNKGMEHLYSMK---CKNVVP---LYDLLLEMIDAHRIJHAP-----	552
CAA05410.2	(LBDG5)	FVMLLKISHIRHMGVVAHCLYDMK-----EKKFVLPSSWKRVEKLGKYVETLRTEKEHRAAEA	610
BAB62888.1	(LBDG11)	---LAASDDYR-----GCL-DLKIEELKSEVLPSSWVEMRKYMETLRTENEHRAAEA	325
AAH03486.1	(LBDG12)	---LAASDDYR-----GCL-DLKIEELKSEVLPSSWVEMRKYMETLRTENEHRAAEA	324
AAK49953.1	(LBDG13)	---LAASDDYR-----GCL-DLKIEELKSEVLPSSWVEMRKYMETLRTENEHRAAEA	323
3ERT		----- 552	
CAA05410.2	(LBDG5)	SPQT 614	
BAB62888.1	(LBDG11)	PPQT 329	
AAH03486.1	(LBDG12)	PPQT 328	
AAK49953.1	(LBDG13)	TPQT 327	

Figure 19 PartII

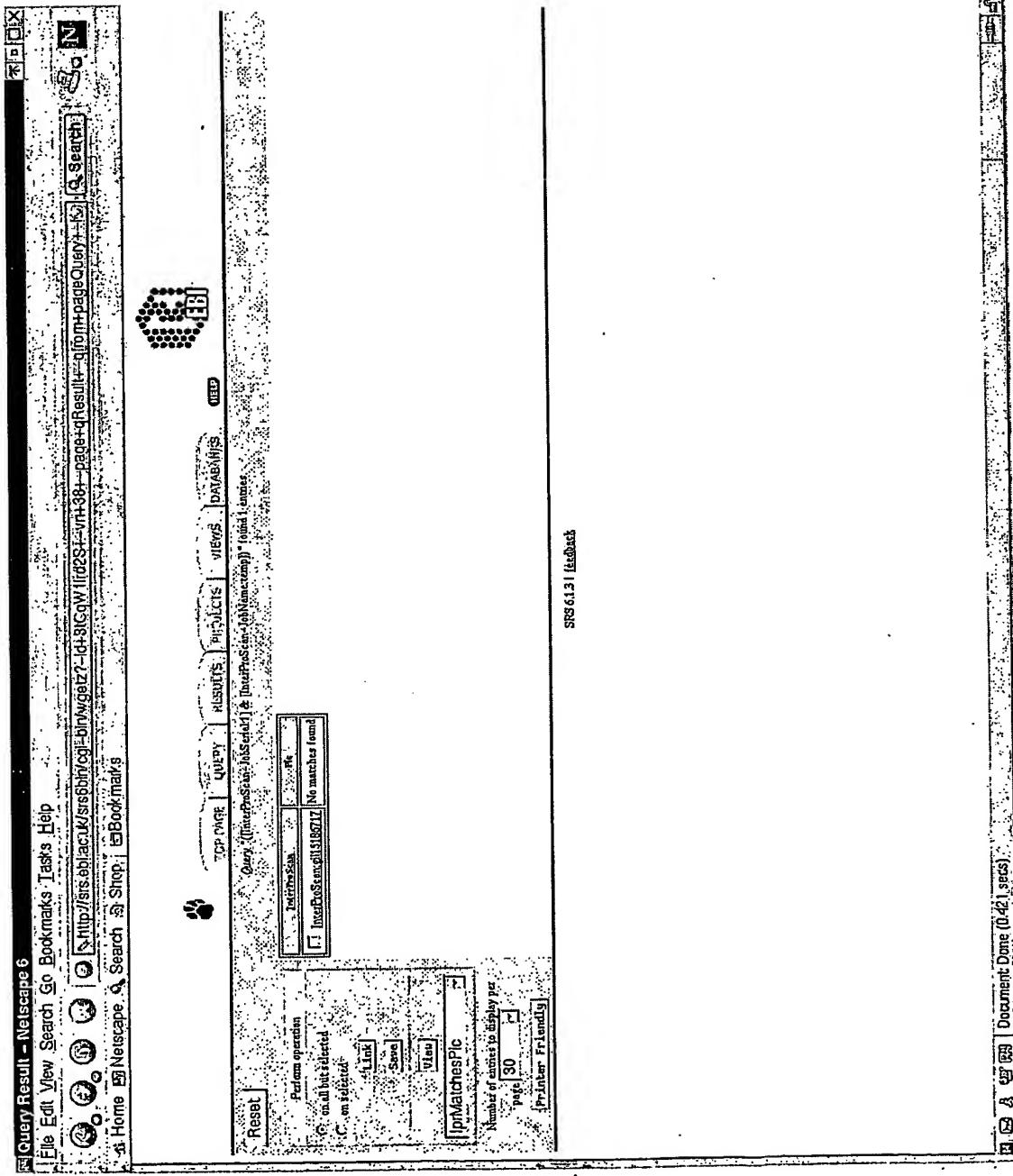
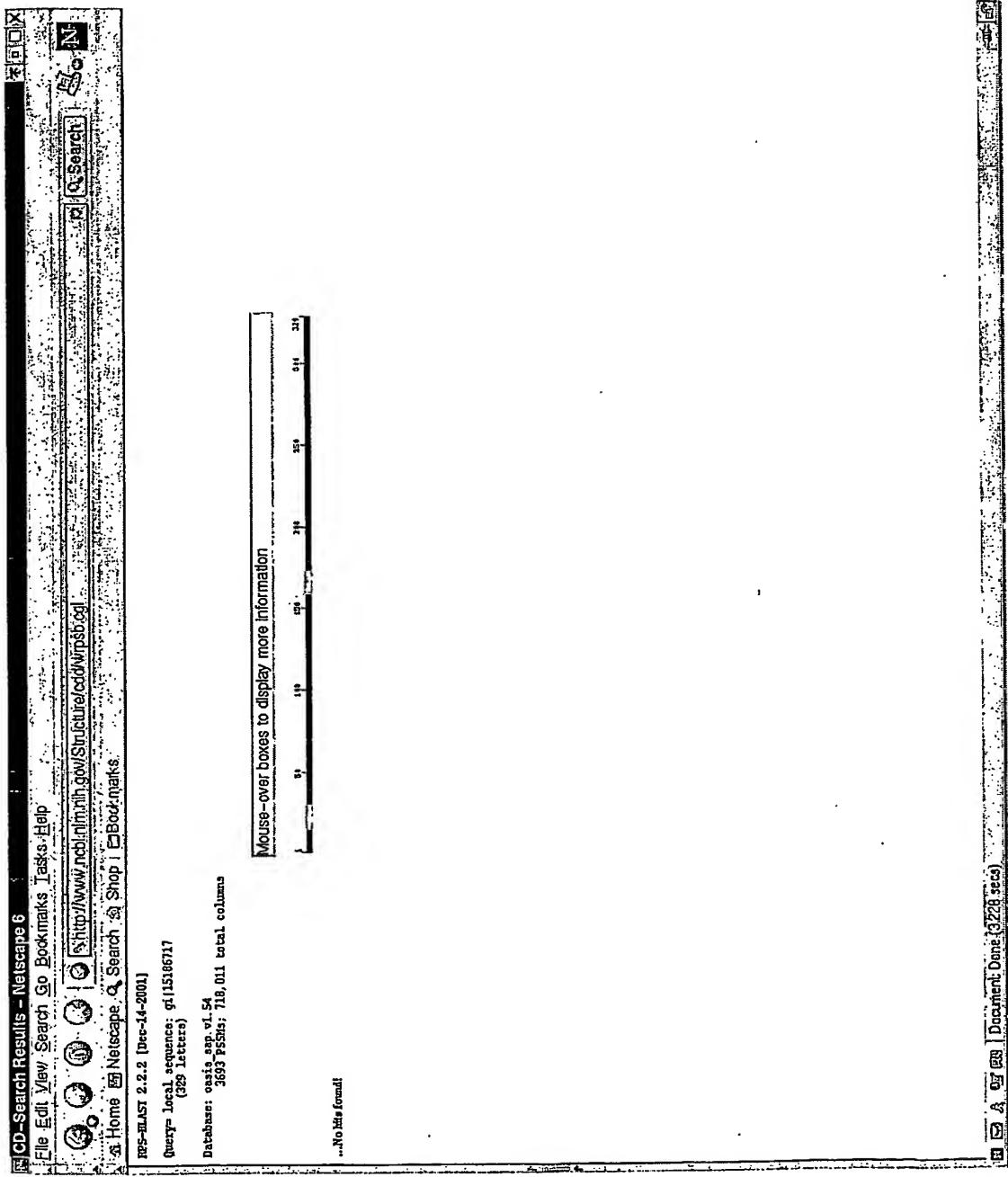


Figure 20



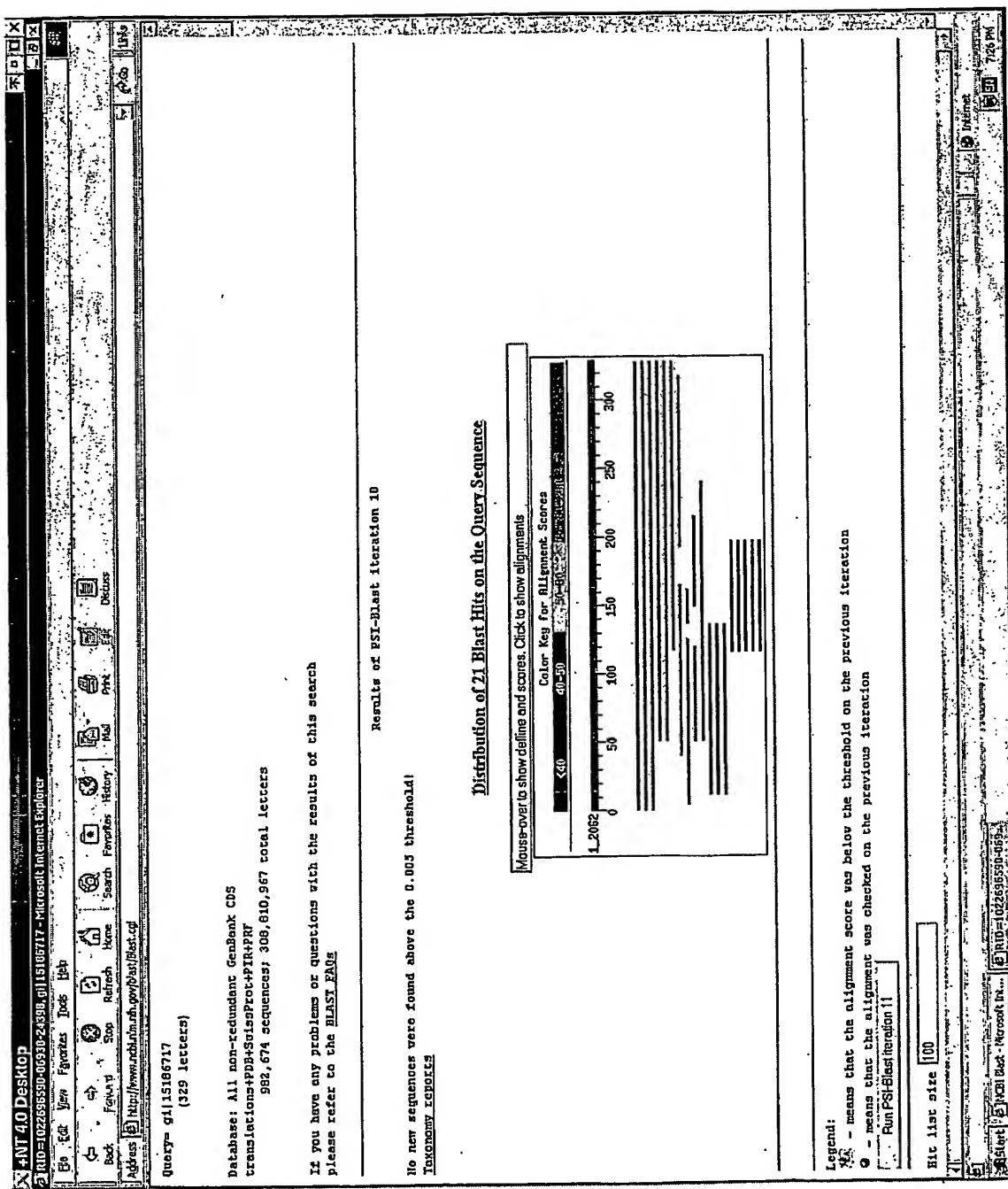


Figure 22

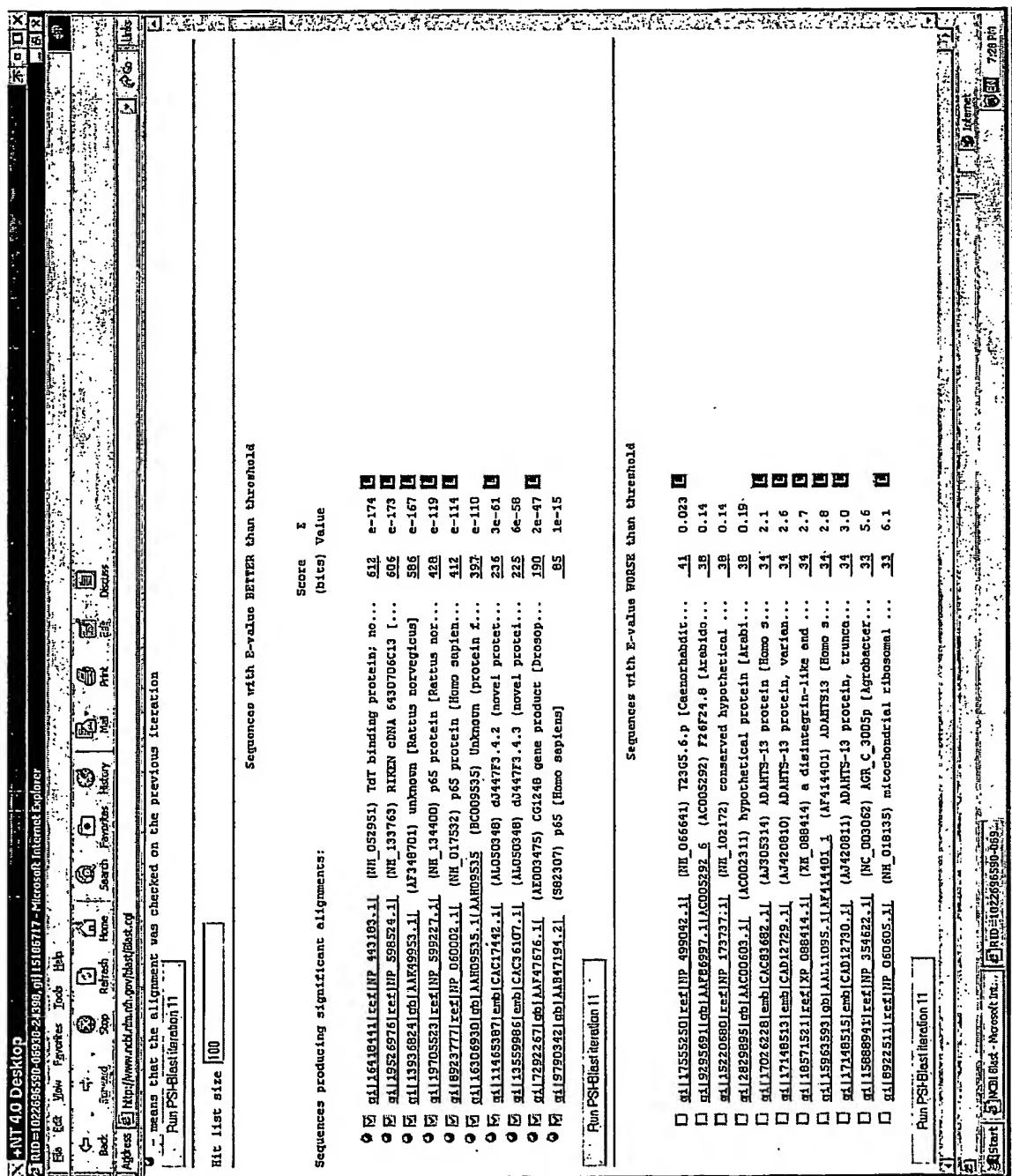


Figure 23

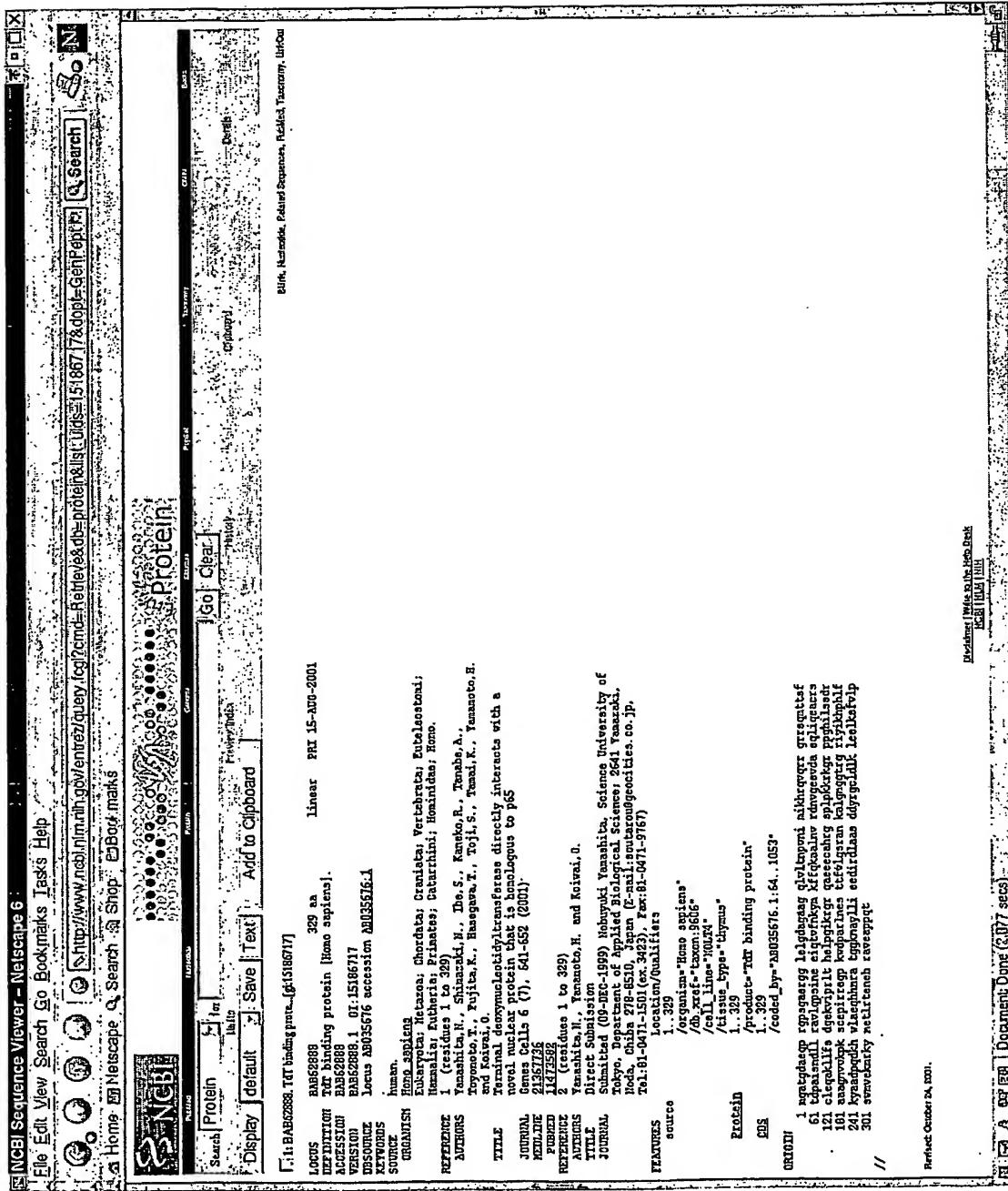


Figure 24

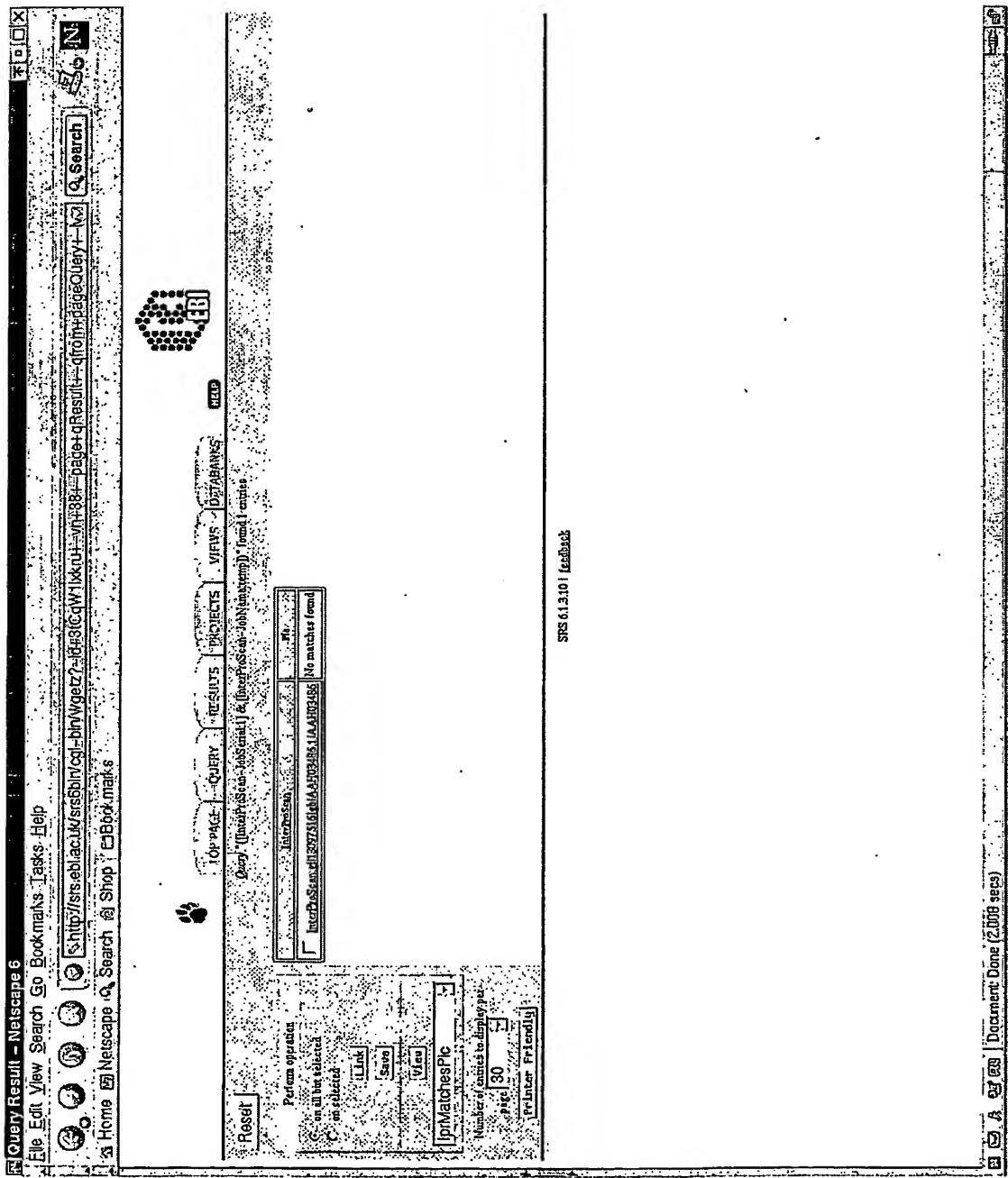
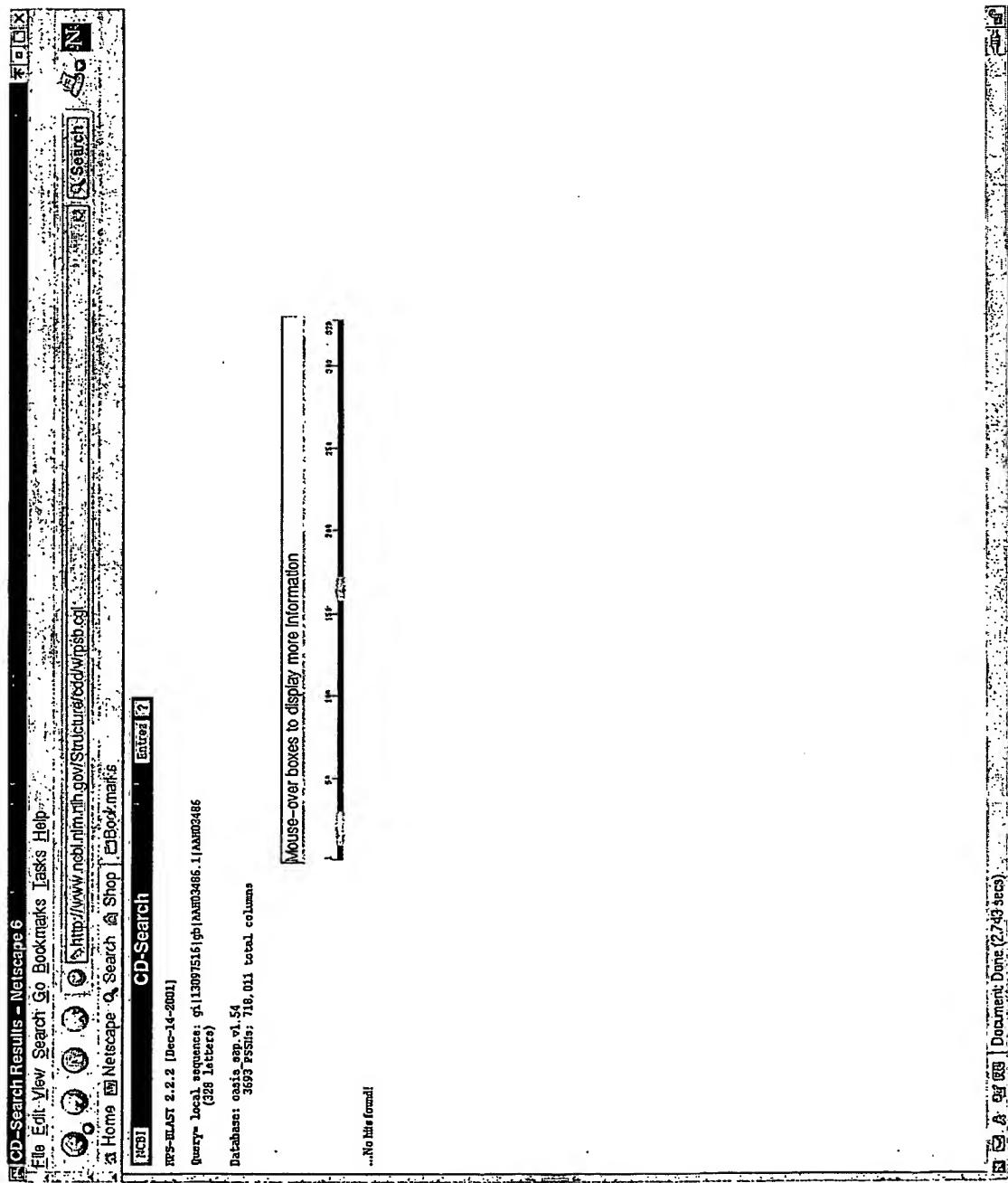


Figure 25



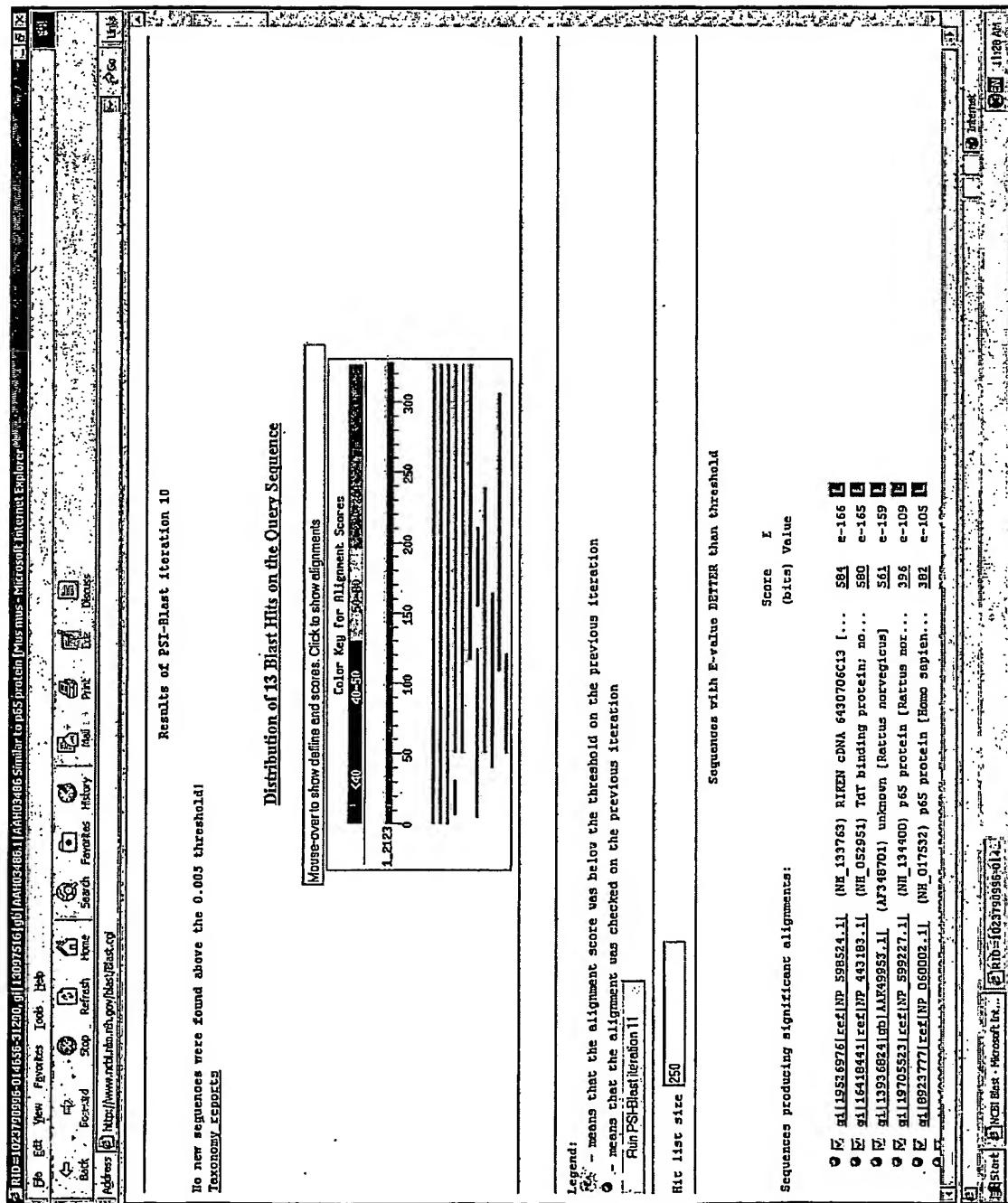


Figure 27

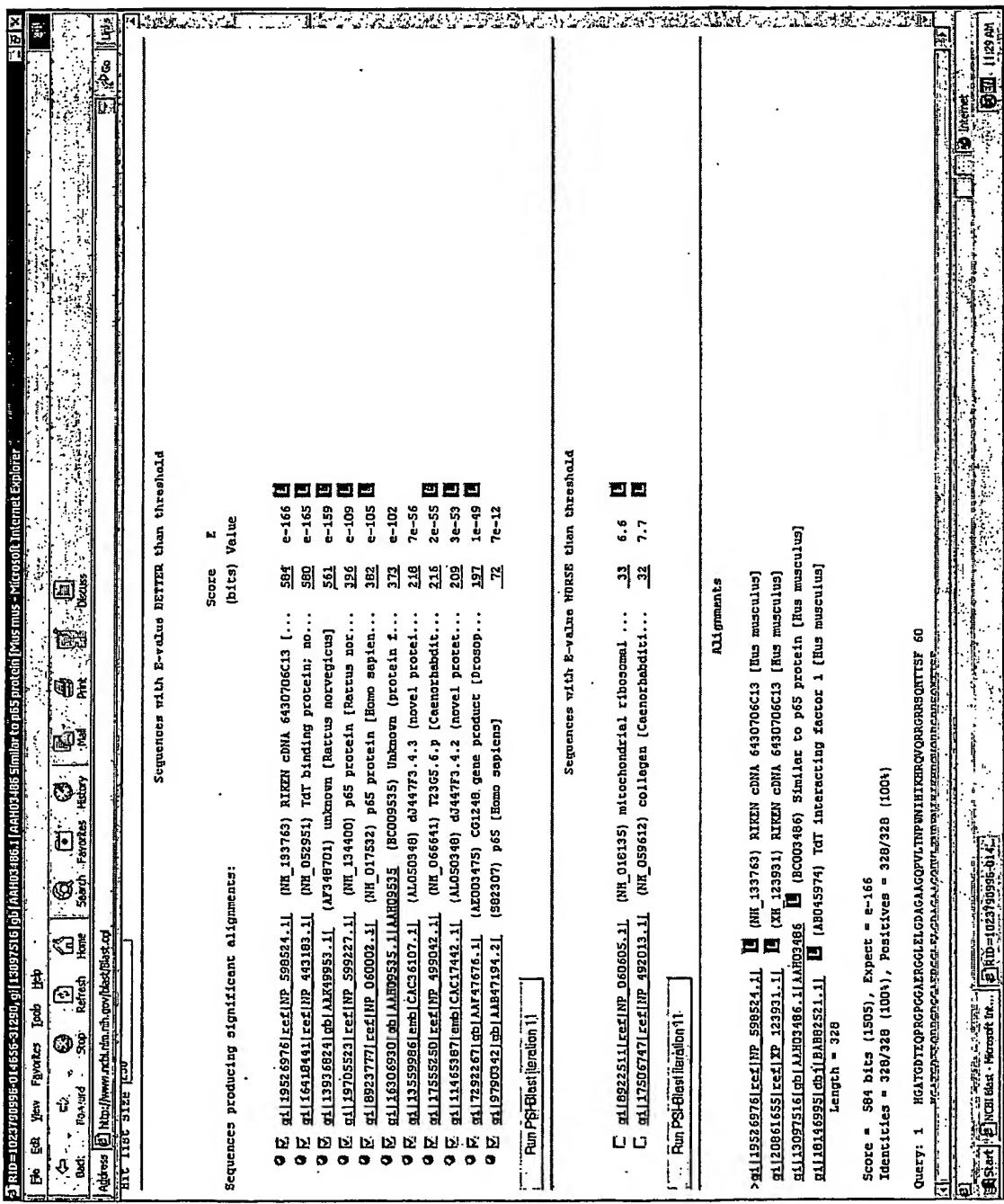


Figure 28

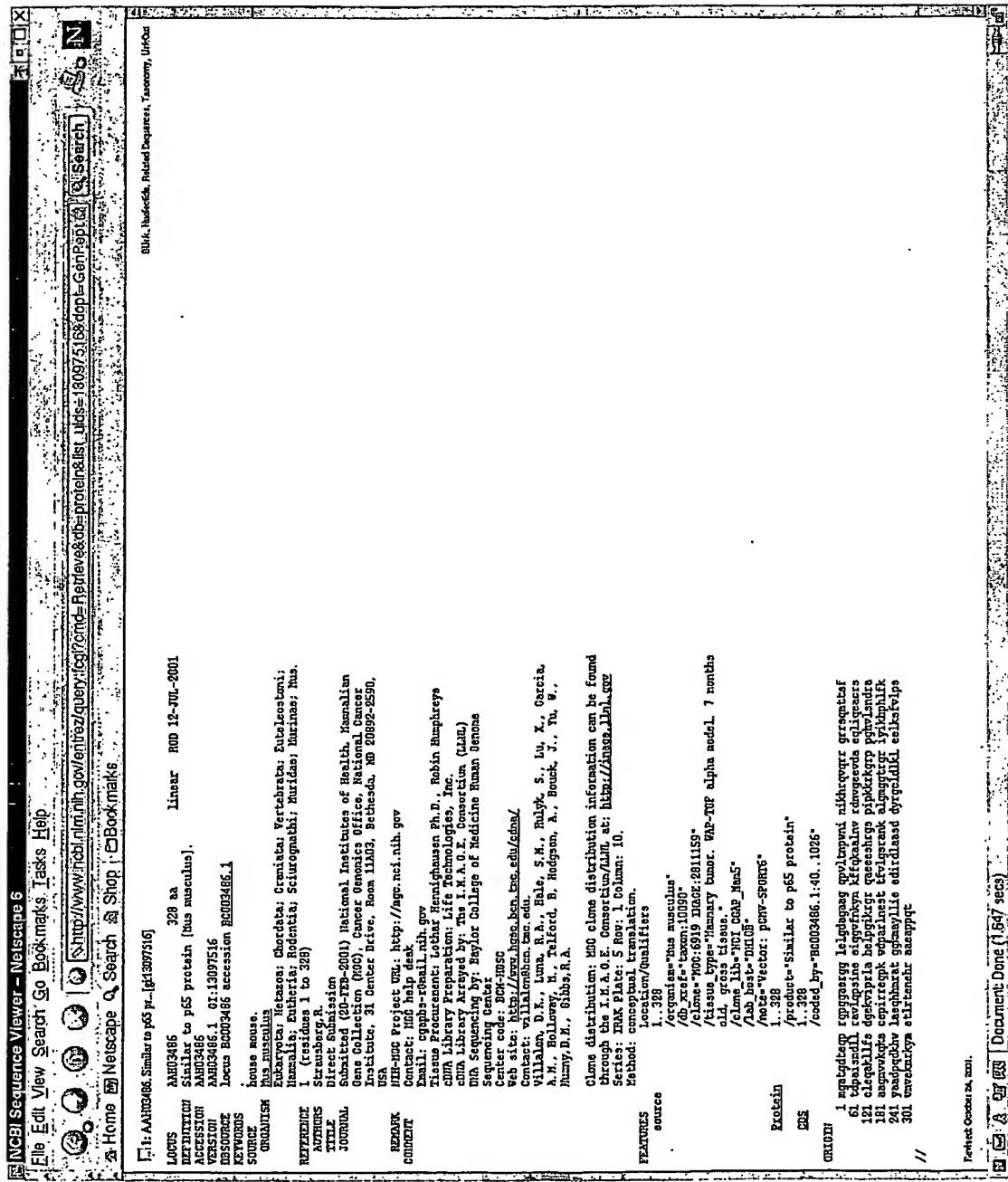


Figure 29

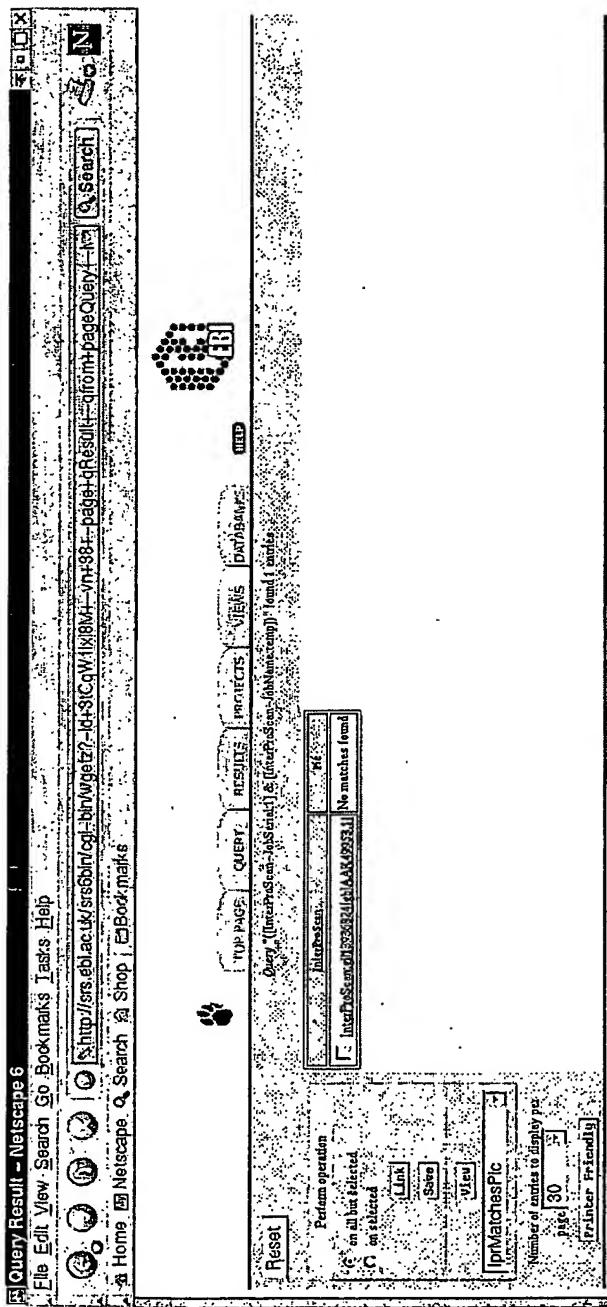


Figure 30

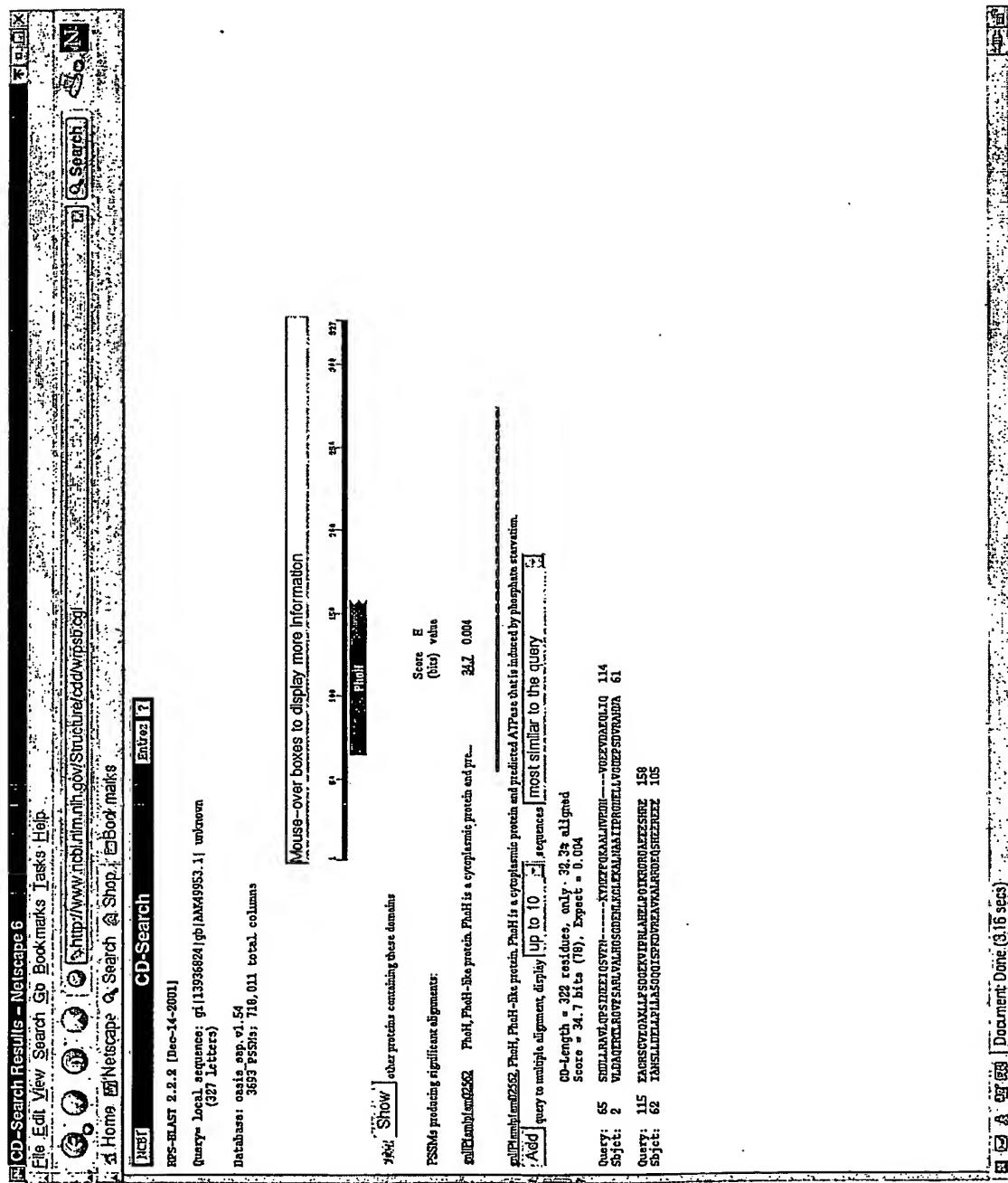


Figure 31

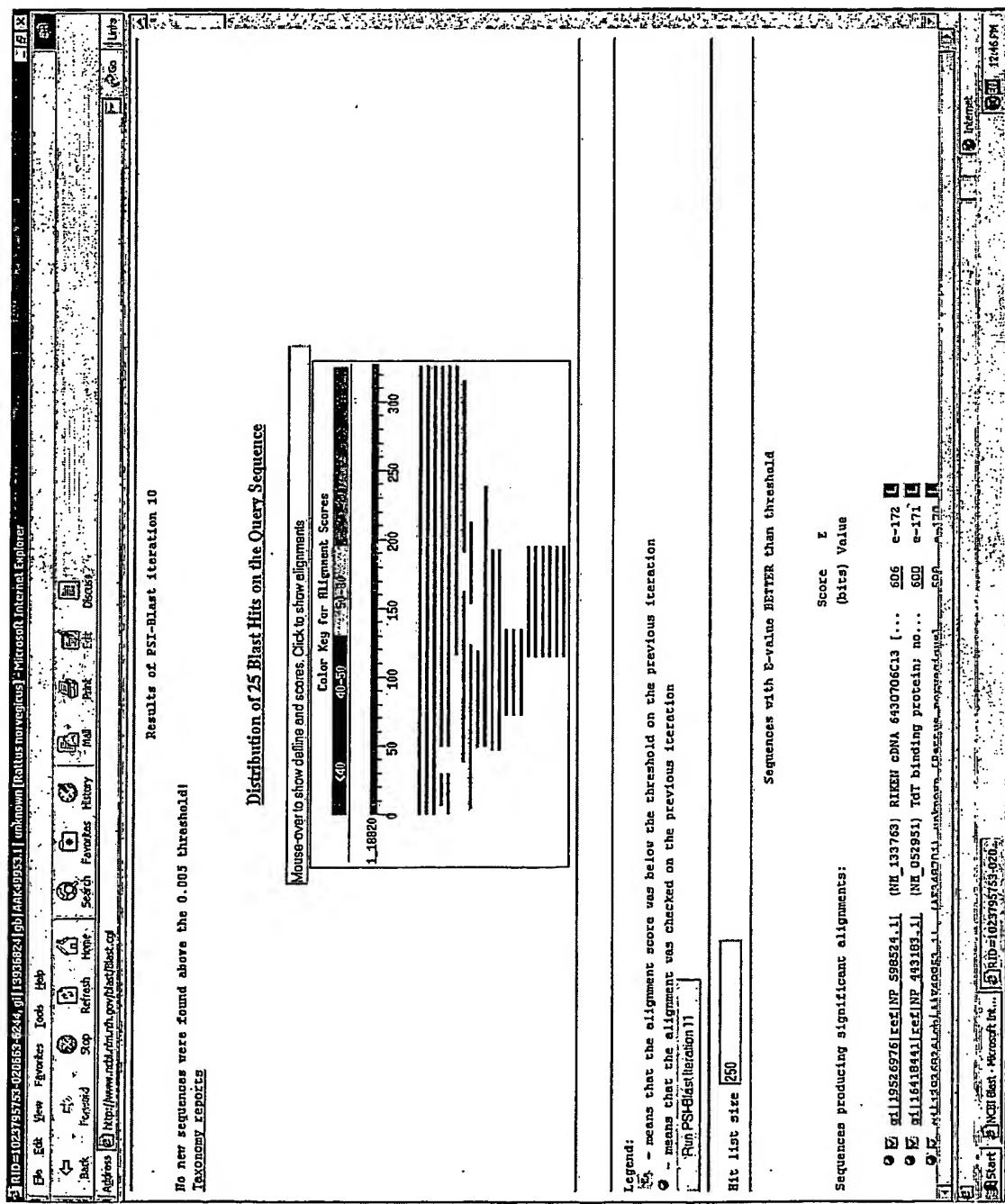


Figure 33 shows a screenshot of a Microsoft Internet Explorer browser window. The address bar shows the URL: <http://www.ncbi.nlm.nih.gov/blast/blast.cgi>. The page displays the results of a BLAST search, specifically the 'Sequences with E-value BETTER than threshold' section.

Sequences producing significant alignments:

	Score	E (bits)	Value
g1.1195269761.ref NP_598524.11 (NH_133763) RIKEH CRNA 6430705613 [...	506	e-172	1
g1.16418441.ref NP_493183.11 (NH_052951) Tdt binding protein; no...	600	e-171	1
g1.139369241.ref NP_147955.11 (AR348701) unknown [Rattus norvegicus]	528	e-170	1
g1.19705523.ref NP_599227.11 (NH_134400) p65 protein [Rattus nor...	434	e-120	1
g1.19237771.ref NP_660002.11 (NH_017532) p65 protein [Homo sapien...	417	e-116	1
g1.16306930.ref NP_149055.11 (NH_009335) Unknown [protein f...	392	e-108	1
g1.1144653871.ref NP_047442.11 (AA050346) dJ44773_4.2 (novel protein)	231	8e-60	1
g1.1355599861.ref NP_3107.11 (AL050348) dJ44773_4.3 (novel protein)	219	6e-56	1
g1.12222671.ref NP_4766.11 (AE004475) CG1248 gene product [Drosophila...	190	2e-47	1
g1.19790342.ref NP_047194.21 (S82307) p65 [Homo sapiens]	85	9e-16	1

Sequences with E-value BETTER than threshold

	Score	E (bits)	Value
g1.17555250.ref NP_499042.11 (NH_066641) T2365_6.3 [Caenorhabditis...	40	0.034	1
g1.20536271.ref XP_032459.21 (XU_092459) K11A1946 protein [Homo sa...	35	1.1	1
g1.1923691.ref NP_1AC058927.11/AC05892.6 (AC005892) R26f724.8 [Arabidop...	35	1.2	1
g1.15220560.ref NP_1723797.11 (NH_102472) conserved hypothetical protein...	35	1.3	1
g1.16916743.ref NP_1BA85532.11 (AB07526) K11A1946 protein [Homo sa...	35	1.3	1
g1.28698851.ref NP_1AC00603.11 (AC00311) hypothetical protein [Arabidop...	35	1.4	1
g1.121265043.ref NP_605955.11 (NH_139026) a disintegrin-like and...	34	1.8	1
g1.121265046.ref NP_605956.11 (NH_139027) a disintegrin-like and...	34	2.0	1
g1.121265013.ref NP_605957.11 (AJ420810) ADAMTS-13 protein, variant...	34	2.0	1
g1.121265034.ref NP_605954.11 (NH_139025) a disintegrin-like end...	34	2.3	1
g1.138571521.ref XP_038444.11 (XH_038444) a disintegrin-like end...	34	2.3	1
g1.121265049.ref NP_605957.11 (NH_139028) a disintegrin-like end...	34	2.7	1
g1.127506747.ref NP_492013.11 (NH_059612) collagen [Caenorhabditis...	34	2.7	1
g1.17555254.ref NP_59274.11 (NH_076873) cuticle collagen [Homo sa...	32	7.3	1
g1.18922511.ref NP_066055.11 (NH_018135) mitochondrial ribosomal ...	32	7.7	1

Run PSI-Blast iteration 11

Figure 33

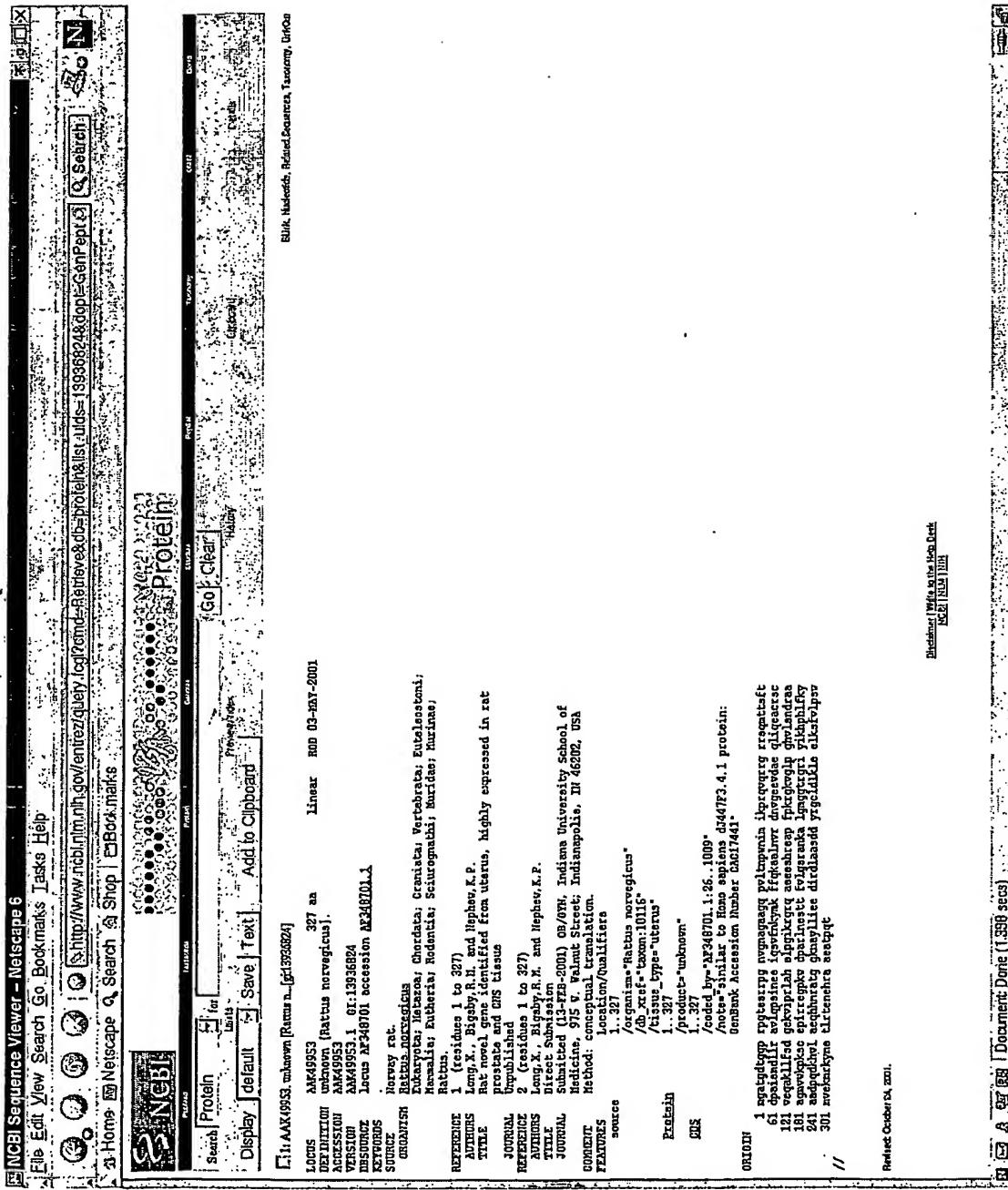


Figure 34



**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.